INSIGHT
Intestinal microbiota

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



SHAPING THE EARTH

The physics of plate tectonics PAGE 140

RESEARCH

HAS THE THESIS HAD ITS DAY?

The past, present and future of the PhD assessment

PAGES 22 & 26

PUBLICATIONS

LET'S BE MORE SCIENTIFIC

Thirty years of honing peer review

PAGE 31

CONSERVATION

THE IDEA OF LANDSCAPE

From 'Capability' Brown to US national parks

PAGE 34

NATURE.COM/NATURE
7 July 2016 £10
Vol. 535, No. 7610

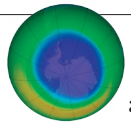
TIFF Preview

THIS WEEK

EDITORIALS

ZIKA Virus is just one of multiple birth defect threats **p.8**

WORLD VIEW Bioweapons treaty must update its science **p.9**



OZONE Earth's radiation shield on the mend again **p.12**

Doctor, doctor

Writing a PhD thesis is a personal and professional milestone for many researchers. But the process needs to change with the times.

According to one of those often-quoted statistics that should be true but probably isn't, the average number of people who read a PhD thesis all the way through is 1.6. And that includes the author. More interesting might be the average number of PhD theses that the typical scientist — and reader of *Nature* — has read from start to finish. Would it reach even that (probably apocryphal) benchmark? What we know for sure is that the reading material keeps on coming, with tens of thousands of new theses typed up each year.

To what end? Reading back over a thesis can be like opening up a teenage diary: a painful reminder of a younger, more naive self. The prose is often rough and rambling, the analyses spotted with errors, the methods soundly eclipsed by modern ones. And students in the process of writing a thesis can find themselves in a very dark place indeed: lost in information, overwhelmed by literature, stuck for the next sentence, seduced by procrastination and wondering why on earth they signed up to this torture at all.

Two News Features this week reflect on that question. They examine the past, present and future of the PhD thesis and the oral examination that often accompanies it. On page 22, three leading scientists — including Francis Collins, director of the US National Institutes of Health — dig out and reread their theses for us, and talk about what they learned. Their musings (filmed and available in a series of videos at go.nature.com/297qrah) show, reassuringly, that they are just the same as the rest of us. They made mistakes, had moments of self-doubt and considered quitting. (Collins actually did quit.) But their stories also reveal how it is important to have the long view in mind.

Thumbing through their theses now, they see how much they learned about the scientific process and how to conduct rigorous research. They realize how precious it was to be able to devote themselves to a single piece of original and creative work. And they feel a sense of accomplishment and pride — as everyone tends to after any difficult life challenge that they struggle with and eventually conquer.

Completing a thesis represents a coming of age not just scientifically, but also educationally and personally. It signals the passing of an intellectual milestone — from a student under the care of a supervisor to an individual who asks questions of their own. It marks the end of formal education, and graduation to a new phase in life. For many people, it also sees their departure from science altogether. Often, the PhD years coincide with significant personal events, as we mature emotionally and meet friends, partners and colleagues who will stay with us for life. All this can also turn thesis-writing into a more significant event than merely the writing up of a (usually) minor piece of science.

Still, it's perhaps too easy to get sentimental over the thesis. For a start, the process has to keep up with the times. The PhD is already assessed in many different ways around the world (as the second News Feature, on page 26, describes) and scientists should welcome ways to keep it relevant. The goal of PhD assessment everywhere remains, rightly, to demonstrate that a student has conducted, and can communicate, independent, original research. But the way in which that's achieved can and should be improved.

For one thing, it doesn't have to involve a vast printed volume. A lot of students could do themselves, their supervisors, their examiners and their wider audience a favour by keeping it crisp and short. Postgraduate supervisors should stress this at the beginning. And it's important to make the work in the thesis available to future researchers by publishing or sharing the data in some form. To contribute to the world beyond the author's immediate circle, a PhD thesis should be read and used, and not just serve as a shelf ornament or doorstop.

"Students could do themselves and their audience a favour by keeping it crisp and short."

For those inspired to go back to their own thesis, and those who are examining a freshly written one, it's best to be kind. As long as the fundamentals are there — the question is interesting and the approach and analysis rigorous — it's fair to forgive the typos and the research paths that turned out to be dead ends. A PhD is, after all, training in research, and to try — and fail — is a valuable part of that course.

Do you know where your PhD thesis is? Dig it out and share with @NatureNews on Twitter using the hashtag #the3wordthesis. You might even bump up that average readership. ■

False assumptions

US regulators must regain the upper hand in the approval system for stem-cell treatments.

You may have heard that regulators in the United States are too strict when it comes to stem-cell treatments. If not, then you will probably hear that message soon — patient groups, entrepreneurs and politicians are broadcasting it as they lobby for a change in the law. The Food and Drug Administration (FDA), this narrative asserts, is holding back effective therapies and, in the words of the most extreme, killing people by blocking their access to cures.

This is false. The claim that regulation is too harsh wrongly implies that the FDA is holding back therapies that work. Critics point to decades of preclinical and clinical work with stem cells and the pipelines of stem-cell treatments. With circular logic, they argue that, because the treatments have not been approved, there is something wrong with the approval system.

The assumption in these accusations — that these treatments work — is at the heart of the problem. The FDA is right to insist that

© 2016 Macmillan Publishers Limited. All rights reserved.

only proper clinical trials can make that case. And the agency's critics are right to point out that this process is lengthy and expensive — perhaps too much so.

The proposed change in the law — the REGROW Act — would tackle this problem by simply doing away with the need for proper trials. It would effectively borrow a fast-track system that Japan created for stem-cell treatments and regenerative medicine. *Nature* has previously expressed concern about this system (see *Nature* **528,** 163–164; 2015). It is not a fit and proper model to export, chiefly because it grants "conditional approval" to treatments with minimal safety data and little attention to efficacy.

Therapies approved under this scheme can be marketed for a given period — around six years — after which time the treatment provider must report back on whether the treatment it has been selling to patients was safe and effective.

In other words, patients (who in Japan have to pay up to 30% of the cost even of treatments covered by national insurance) are subsidizing clinical trials. Most of these treatments, as the history of phase III trials shows, will probably fail. People who took an ineffective drug (and probably spurned other treatments to do so) will not get their money back.

Japan still has to prove that data collection under this system will be rigorous enough to prove a treatment's efficacy. And if the system works and drugs are found to be ineffective, the regulatory agency will then have to fight the uphill battle of reining back treatments that were already on the market but are now de-approved.

Overall, Japan will most probably see a flood of safe but ineffective treatments. That scenario would discourage anyone from going through the costly steps required to create therapies that really do work (if you can sell garbage for the same price, why not stick with that?). That would be a shame for a field with such promise. Is this the way the United States wants to go?

Another reason for saying that the FDA is not unduly harsh on restricting stem cells is the large number of clinics that already operate and sell unapproved treatments. A study released last week reported 351 businesses offering stem-cell treatments at 570 clinics in the United States (L. Turner and P. Knoepfler *Cell Stem Cell* http://doi.org/bkpv; 2016). Although the study does not directly accuse these clinics or businesses of wrongdoing, many of them promise stem-cell treatments for neurodegenerative diseases for which no stem-cell treatment has so far proved effective.

"The assumption that these treatments work is at the heart of the problem."

These treatments, which usually claim that a certain type of stem cell can transform into another type of mature cell able to ameliorate such diseases, require approval by the FDA. The existence of these clinics shows that the FDA is not strict — never mind too strict — in its regulation.

That the FDA moves so slowly to crack down on existing unapproved stem-cell treatments makes the prospect of conditional approval — an opportunity to embed ineffective treatments in the US health-care system — all the more worrisome.

The best way for the FDA to respond to the mood that has seeded the REGROW Act is to agree on a more efficient way to approve cell treatments. It is working to do so, but tensions are high. A hearing planned for April was overwhelmed by prospective participants. It is now scheduled for September — stretched to two days and with a public workshop added.

The FDA should strive to keep this debate on the proper topic — how to create a more efficient system that still scientifically evaluates whether treatments are safe and efficacious. To fall short would be a setback for science, and for patients. ■

Beyond Zika

The spotlight on Zika virus should help to spur broader research into birth defects.

In the time it takes you to read this article, a baby will be delivered in the United States with a birth defect. That's about 120,000 every year. For the many individuals with severe cases, childhood and beyond becomes a struggle with mental or physical disabilities, hospital visits and day-to-day worries. And that is in one of the world's richest countries. In low- or middle-income countries, surveillance of birth defects is often absent or so weak that health authorities simply don't know the scale of the problem, making it difficult to develop appropriate prevention measures and care.

The harsh realities of birth defects are shown in recent photographs of babies born in Brazil with abnormally small heads — a condition called microcephaly that seems to be linked to the mosquito-borne disease Zika. The threat of the Zika virus has put birth defects on the political and public-health agenda in a way not seen since the rubella virus (the cause of German measles) led to a pandemic of such defects in the mid-1960s.

Zika therefore provides an opportunity to greatly raise awareness of birth defects, and to bolster support for research and improved public-health action on their many other preventable causes. Researchers must urgently make this case to funders and their political paymasters before the flurry over Zika inevitably ebbs (see page 17).

One target should be the eradication of rubella. It is a scandal that, worldwide, some 100,000 babies are born annually with congenital rubella, despite the availability of a cheap and effective vaccine. The virus spreads slowly and is a low-hanging fruit for eradication through accelerating vaccination in poorer countries.

Another easy target is the compulsory addition of folate vitamins to food staples to protect against neural-tube defects, such as spinal bifida, in developing fetuses. Despite a wealth of evidence that compulsory fortification works, as well as its adoption in the United States, most countries (including all European ones) have yet to follow suit.

The longer-term challenge is to develop the research infrastructure needed to find and prevent the causes of birth defects — in particular because a whopping three-quarters of occurrences have no identified cause. Some will prove to be random events, and others will have genetic or multifactorial origins, but it is likely that many are down to environmental or infectious exposures that public-health authorities can do something about.

This sort of research requires long-term commitment and investment, and the nurturing of highly specialized research communities. Of all the types of epidemiological research, studies of birth defects are perhaps the most difficult. Although their combined human and public-health impact is enormous, individual congenital abnormalities are relatively rare in comparison with, say, lung disease. This means that population-scale databases are needed to capture and record birth defects, and to achieve adequate statistical power.

Amid the political climate of Brexit, there is a certain irony that one of the most developed surveillance systems for birth defects, the European Surveillance of Congenital Anomalies (EUROCAT), was conceived with far-sighted vision in 1974 by the then European Economic Community in the wake of the tragedies of rubella and the drug thalidomide. Such registries may seem mundane, but they are crucial if we are to underpin exploration of the causes and risk factors of congenital anomalies and to provide an early-warning system for new causes of birth defects.

Birth defects should be a top public-health priority to protect the youngest and most vulnerable members of our society. It is staggering in 2016 that they are not. ■

© 2016 Macmillan Publishers Limited. All rights reserved.



UNIV. BRADFORD

Find the time to discuss new bioweapons

The Biological Weapons Convention needs to take the assessment of emerging scientific dangers more seriously, argues **Malcolm Dando**.

In Geneva next month, officials will discuss updates to the global treaty that outlaws the use of biological weapons. The 1972 Biological Weapons Convention (BWC) was the first agreement to ban an entire class of weapons, and it remains a crucial instrument to stop scientific research on viruses, bacteria and toxins from being diverted into military programmes.

The BWC is the best route to ensure that nations take the biological-weapons threat seriously. Most countries have struggled to develop and introduce strong and effective national programmes — witness the difficulty the United States had in agreeing what oversight system should be applied to gain-of-function experiments that created more-dangerous lab-grown versions of common pathogens.

As scientific work advances — the CRISPR gene-editing system has been flagged as the latest example of possible dual-use technology — this treaty needs to be regularly updated. This is especially important because it has no formal verification system. Proposals for declarations, monitoring visits and inspections were vetoed by the United States in 2001, on the grounds that such verification threatened national security and confidential business information.

The treaty therefore relies on countries converting its prohibitions into national law, and setting up proper regulations and oversight. But there is a problem with the way that the BWC is set up to receive and process scientific advice, which affects the ability to update it efficiently. Next month's meeting must address this problem, and scientists who care about the societal impacts of research should lobby their elected politicians to make sure that it does.

The BWC is formally reviewed every five years at a special conference (the next is in November this year, and the Swiss August meeting in is to prepare for it). During the intervening years, annual one-week meetings of government experts, and later of government representatives (state parties) are intended to track progress and raise issues. But there is not enough time at these meetings to discuss what is needed in sufficient depth. So no properly thought-out recommendations can be made.

In 2013, for instance, the experts' meeting scheduled a mere six hours of discussions on science and technology — less than a day. That is not enough time for complex science to be presented, digested and discussed, and not enough to consider its implications and suggest revisions to the BWC.

There is widespread awareness that the current system is not fit for purpose. At a preparatory meeting in April, 5 of the 13 working papers dealt with the need to find a better way to carry out these crucial interim discussions on science and technology. As the Russian paper noted: "There is widespread agreement that improved and more effective arrangements are required."

Other international agreements have effective ways to track and deal with scientific and technological change. The 1997 Chemical Weapons Convention has a permanent scientific advisory board. When concerns were raised in 2011 about the possible harmful implications of the convergence of chemistry and biology, that board set up a dedicated working group to investigate and report back. It did so in 2014 — concluding that the current threat was low but that future developments should be monitored closely. The assessment system led to action. At present, the BWC assessment system cannot.

In the long term, the BWC may need a similar advisory board for science. But that is unlikely to happen soon, and as science is rapidly changing, we have to find a way to improve the way the interim annual meetings work. My colleague Kathryn Nixdorff and I interviewed delegates at previous meetings about possible improvements, and we have some simple suggestions.

STRONG INTERNATIONAL ACTION IS NEEDED TO ASSESS THE THREATS FROM THE NEW AGE OF BIOLOGICAL TECHNIQUES.

The discussions of science at the experts' meetings should be split off into a separate dedicated parallel track. This is the best way to create the necessary time. Even then, it will be impractical to cover all relevant ground across the sciences, so each year a specific topic — CRISPR editing, say — should be considered. Researchers and scientific bodies should present the facts, and then discuss the implications with government officials at the experts' meeting. Who should attend these sessions? We argue that they should be open to representatives from any member state.

Feeding back results of these expert discussions to the broader BWC, a designated diplomat — in place for the full five-year period between review conferences — would attend the annual experts' meetings and write a report. The annual meetings of state parties should then assess these reports and agree any action needed. Future review conferences should check on progress.

Even so, issues such as the possible dual-use threat from gene-editing systems will not be easily resolved. But we have to try. Without the involvement of the BWC, codes of conduct and oversight systems set up at national level are unlikely to be effective. The stakes are high, and after years of fumbling, we need strong international action to monitor and assess the threats from the new age of biological techniques.

If the BWC cannot find a way to adapt to the pace of scientific and technological change, then it risks becoming irrelevant as the world searches for biosecurity in the coming decades. ■

Malcolm Dando *is professor of international security at the University of Bradford, UK.*
e-mail: m.r.dando@bradford.ac.uk

© 2016 Macmillan Publishers Limited. All rights reserved.

RESEARCH HIGHLIGHTS

Selections from the scientific literature

NEUROIMMUNOLOGY

Reward system boosts immunity

Activating the reward system in the brains of mice directly boosts their immune systems, offering a physiological explanation for the placebo effect.

Shai Shen-Orr, Asya Rolls and their colleagues at the Technion–Israel Institute of Technology in Haifa activated neurons in a part of the mouse brain that processes rewarding activities such as eating and sex. The next day, they injected the mice with the bacterium *Escherichia coli*. The animals showed increases in both short-term and long-term immune responses to the pathogen, compared with mice in a control group. But these effects were lost when the researchers also inactivated the animals' sympathetic nervous systems, suggesting that this system helps to mediate interactions between the brain and the immune system.

Nature Med. **http://dx.doi.org/10.1038/nm.4133 (2016)**

ASTROPHYSICS

No neutrinos from black hole smash

The first hunt for neutrinos coming from the merger of two black holes — which last year produced the first direct detection of gravitational waves — has come up empty.

Imre Bartos at Columbia University in New York and his colleagues analysed data from two neutrino detectors: ANTARES, under the Mediterranean Sea, and IceCube at the South Pole. They found that no neutrinos were detected at ANTARES in the 500 seconds before or after the black holes collided, and that just three were detected at IceCube — none of which came from the direction of the event.

The scarcity of neutrinos from the collision puts an upper limit on how much energy it could have radiated through the near-massless particles, say the authors. If researchers can find a signal from a black-hole collision in the future, they could use the relatively high spatial resolution of neutrino telescopes to pinpoint its location.

Phys. Rev. D 93, **122010 (2016)**



H. WEIMERSKIRCH/CEBC, CNRS

ZOOLOGY

Wind powers weeks of non-stop flight

Frigate birds use the power of the wind and rising air to stay airborne for many weeks at a time.

Henri Weimerskirch at the CNRS Centre for Biological Studies in Chizé, France, and his colleagues fitted great frigate birds (*Fregata minor*), with devices to track their movements over the Indian Ocean. Some birds were also fitted with devices to measure their heart rate and acceleration.

The researchers found that the birds stayed on the wing for up to 48 days and travelled an average of 450 kilometres daily, often tracking the wind around the edge of the huge area of low pressure called the doldrums.

The birds use a "roller-coaster flight", the authors say, ascending up to 4,000 metres with the help of the wind and thermals. Frigates cannot land on the water, but they can glide over distances of many kilometres in a low-energy mode — sometimes with no flapping at all. This may provide them with the opportunity to nap for up to 12 minutes at a time, and allow them to stay in the air almost indefinitely.

Science 353, **74–78 (2016)**

EVOLUTION

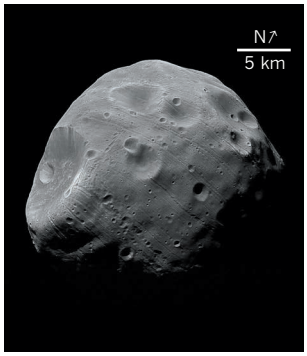
Lizards tailor tails to local predators

Brightly coloured tails are a common feature of young lizards, and can be tailored to the eyesight of specific local predators.

Takeo Kuriyama and his colleagues at Toho University in Funabashi, Japan, collected 15 juvenile *Plestiodon latiscutatus* lizards from three areas of Japan dominated by different predators — snakes, weasels or birds. Lizards' tails were vivid blue where weasels or snakes were common, but had high ultraviolet reflectance only in areas high in snakes. Weasels can see blue wavelengths, but, unlike snakes, cannot detect UV light, suggesting that the lizards have evolved to draw the attention of specific local predator species away from their bodies and towards their disposable tails. Brown tails were found in the area where keen-eyed predatory birds make camouflage a better strategy.

J. Zool. **http://doi.org/bkqm (2016)**

© 2016 Macmillan Publishers Limited. All rights reserved.



ESA/DLR/FU BERLIN (G. NEUKUM)

PLANETARY SCIENCE

Martian moons formed *in situ*

The moons of Mars may have formed from a disk of debris kicked up by the impact of a giant meteorite on the planet.

Astronomers have struggled to explain the existence of Phobos (**pictured**) and Deimos, the small, irregularly shaped moons of the red planet. One view is that they were asteroids captured by Mars. But a team led by Pascal Rosenblatt at the Royal Observatory of Belgium in Brussels tested an alternative idea using computer simulations of how orbiting debris, created by a giant impact, might coalesce.

Relatively large moons form in the inner part of the disk thrown up by such a smash, and migrate outward, causing the outer part of the disk to coalesce into two bodies the sizes of Phobos and Deimos. The inner large moons are eventually dragged inward and fall back to Mars over 5 million years.

Nature Geosci. **http://dx.doi.org/10.1038/ngeo2742 (2016)**

CLIMATE SCIENCE

Warming shifts plant sex ratio

Climate change seems to be skewing the sex ratios of an alpine herb towards male plants.

William Petry at the University of California, Irvine, and his colleagues analysed data on populations of the herb valerian (*Valeriana edulis*) in the Rocky Mountains of Colorado as the region became warmer and drier over the past few decades. They found that in 2011, plants at the highest elevations were only 23% male, whereas at lower altitudes, where the climate is warmer and wetter, the plants were up to 50% male. Across 9 populations at a variety of elevations, there was an average of 6% more males in 2011 than in 1978.

A higher male-to-female ratio could result in increased pollination — and therefore seed production — which could help the plants to expand their range as they adapt to climate change, the authors suggest.

Science 353, **69–71 (2016)**

ROBOTICS

Soft wheels make robots tough

Wheels built entirely from soft materials can help robots to roll over tricky terrain and resist damage.

Aaron Mazzeo and his co-workers at Rutgers University in Piscataway, New Jersey, built a squishy wheel inspired by the inching motions of soft creatures such as earthworms. A stretchable ring contains multiple internal chambers, groups of which can be inflated and deflated sequentially around the circle. The pressurized compartments exert torque on a second, outer ring, causing it to turn.

A soft robotic vehicle fitted with four of these wheels (**pictured**) travelled on a flat surface at 3.7 centimetres per second and kept moving after being dropped from eight times its height. The researchers also drove the robot over a rocky landscape and underwater, and show that their concept can be modified to make winch rotors.

Adv. Mater. **http://doi.org/f3qjsh (2016)**



WILEY

SOCIAL SELECTION

Fake article webpages draw fire

A debate is swirling around a tactic that academic publisher John Wiley & Sons uses to fight online piracy (see go.nature.com/299xi1y). The company created a webpage, accessible by several URLs, that looked like an academic paper to automated downloading programs. But any users who accessed the URLs were then blocked from viewing other Wiley content. Wiley and other publishers use these 'trap' URLs — which are invisible to most human users — to detect and prevent unauthorized downloading and republishing of their content. But some users say that the tactic is too heavy-handed.

NATURE.COM
For more on popular papers:
go.nature.com/29hbqog

CANCER BIOLOGY

Leukaemia cells hide in fat tissue

Cancer-causing stem cells evade chemotherapy by surviving in fat deposits around gonads.

Fat tissue supports the growth of normal blood-forming stem cells. Craig Jordan of the University of Colorado Denver and his colleagues found that in a mouse model of one form of leukaemia, gonadal fat deposits contained numerous cancer stem cells, but subcutaneous fat had very few. Leukaemic cells induced the breakdown of gonadal fat, releasing nutrients that fuelled the growth of malignant cells in fat as well as other tissues. The cancer stem cells also expressed CD36, a cellular marker that boosts fat metabolism, helping to protect the cells from many chemotherapy drugs.

Targeting fat metabolism could help to eradicate leukaemia stem cells, the authors suggest.

Cell Stem Cell **http://doi.org/bkqj (2016)**

CLIMATE CHANGE

Negative carbon emissions needed

Countries' existing promises regarding emissions reductions are unlikely to prevent global warming exceeding 2 °C above pre-industrial temperatures by the end of the century, meaning that large amounts of carbon may need to be removed from the atmosphere.

Benjamin Sanderson and his co-workers at the US National Center for Atmospheric Research in Boulder, Colorado, explored the odds of staying below 2 °C of warming for a range of emissions pathways. They also analysed whether 'negative emissions' — the removal of carbon from the atmosphere — will be necessary.

To avoid crossing the 2-degree threshold during this century, net global emissions must drop to zero by 2085, the authors find. Depending on the level of near-term reductions, between 1.5 billion and 5 billion tonnes of CO_2 per year will need to be captured and removed from the atmosphere thereafter.

Geophys. Res. Lett. **http://doi.org/bkqh (2016)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

© 2016 Macmillan Publishers Limited. All rights reserved.

SEVEN DAYS *The news in brief*

BUSINESS

Volkswagen to pay

Volkswagen has made a US$14.7-billion settlement with regulators in the United States over its manipulation of emissions tests for the diesel engines used in almost half a million cars on US roads. The world's biggest carmaker announced on 28 June that it would set aside $10 billion to buy back or terminate the leases on affected Volkswagen and Audi models made from 2009 to 2015, or to modify the vehicles to reduce their emissions. It agreed to spend $2.7 billion on cleaning up environmental pollution and to invest an extra $2 billion in clean car technology in the United States over the next 10 years.

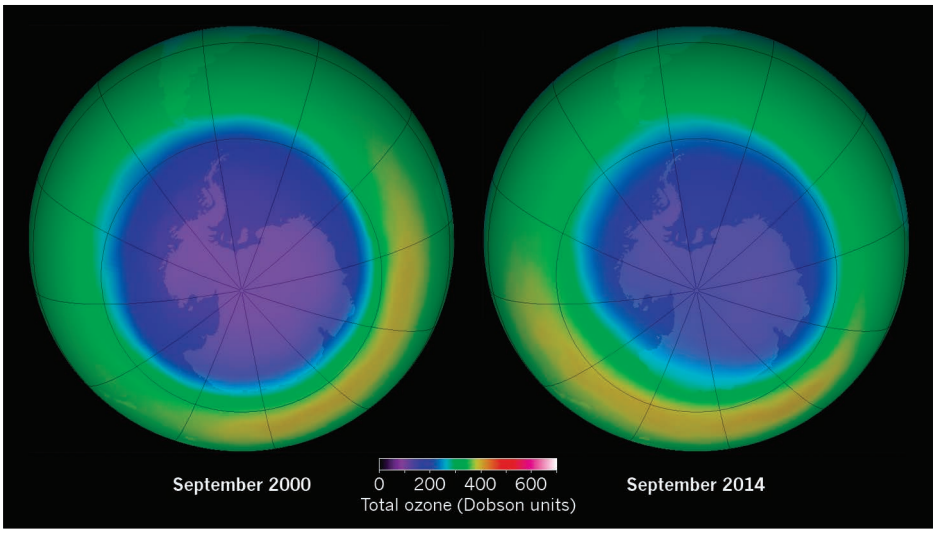
EVENTS

Laureates on GM

More than 100 Nobel laureates have signed an open letter urging environmental group Greenpeace to stop campaigning against genetically modified (GM) organisms. "Opposition based on emotion and dogma contradicted by data must be stopped," they say. The 29 June letter cites, in particular, campaigns against the vitamin-A-enriched 'golden rice', which the laureates say could reduce disease-causing deficiencies of the nutrient. Greenpeace responded on 30 June, saying that golden rice had failed as a solution and suggesting that malnutrition should be addressed through "diverse diet, equitable access to food and eco-agriculture".

Brexit fallout

Scientists in the United Kingdom have continued to voice concerns about the impact of the country's 23 June vote to leave the European Union, amid fears over access to EU funding and the employment rights of academics. High-profile figures including Paul Nurse, director of the Francis Crick Institute in London, have demanded that science be given a seat at the table during exit negotiations. An inquiry into the implications of the vote has been started by politicians from the House of Commons Science and Technology Committee. And science minister Jo Johnson attempted to reassure researchers in a speech in London on 30 June about ongoing plans to reform the country's research-funding system. He noted that the United Kingdom is still open to all EU researchers. See go.nature.com/29v5m8n for more.



NASA OZONE WATCH

Ozone hole shows signs of healing

The Antarctic ozone hole is on the mend, according to an analysis published on 30 June (S. Solomon *et al. Science* http://doi.org/bkn5; 2016). The hole — which opens in the stratosphere every Antarctic spring — has shrunk in its September extent (**pictured**) by an average of 4.5 million square kilometres since 2000. The work used balloon observations and climate models to confirm that ozone is recovering thanks to the 1987 Montreal Protocol, which banned ozone-depleting chlorofluorocarbon chemicals. Fingerprints of healing are most obvious in September, the month the hole begins to grow in earnest. An exception to the healing trend occurred in 2015, when a record hole opened because of a volcanic eruption in Chile. Sulfur particles from the event temporarily accelerated the ozone-destroying reactions. See go.nature.com/29ewicm for more.

Chemist killed

An 'anarcho-primitivist' group has claimed responsibility for the killing of José Jaime Barrera Moreno, a chemist at the National Autonomous University of Mexico (UNAM) who was stabbed to death at the university in Mexico City on 27 June. In a message posted online on 29 June, a group calling itself Individuals Tending Towards Savagery (ITS) says that it carried out the attack. ITS is an alliance of eco-extremist groups that also claimed the 2011 shooting of another UNAM scientist, biotechnologist Ernesto Méndez Salinas. The group has repeatedly attacked scientists and technologists, whom it blames for destroying nature.

RESEARCH

Dawn stays put

On 1 July, NASA turned down a request to send its Dawn spacecraft, currently orbiting the asteroid Ceres, to another destination. Team leaders had wanted to send the mission, which has been at Ceres since March 2015, to the asteroid Adeona. NASA instead opted to keep Dawn where it is, to study Ceres as it gets closer to the Sun. The agency approved mission extensions for all of the

© 2016 Macmillan Publishers Limited. All rights reserved.

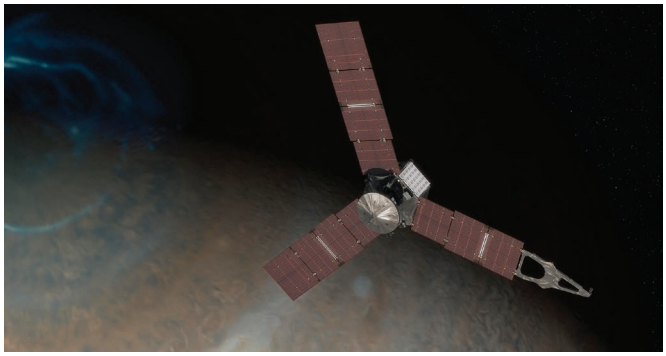
planetary spacecraft that it was considering them for, including New Horizons, which after a successful visit to Pluto last year is now on its way to 2014 MU69, an object in the Kuiper belt. It is expected to fly past the icy world in January 2019.

Rosetta's final move

The Rosetta spacecraft will crash land on comet 67P/Churyumov–Gerasimenko on 30 September, two years after reaching it, the European Space Agency announced on 30 June. In August, engineers will put Rosetta in a series of flattening elliptical orbits, ahead of a final manoeuvre 12 hours before impact. In its final descent, Rosetta will get its closest look yet at the comet. Despite a relatively soft landing, at a speed of 50 centimetres per second, the impact will bring the mission to a close: Rosetta's transmitter will be switched off and its antenna will no longer be able to point to Earth.

Juno finds Jupiter

NASA's Juno spacecraft slipped into orbit around Jupiter on 5 July, becoming the first craft to visit the giant planet since the Galileo mission arrived in 1995. The probe (artist's impression, **pictured**) fired its main engine flawlessly in a 35-minute burn that sent it looping around the planet into a 53.5-day orbit; later this period will be reduced to



NASA/JPL-CALTECH

14 days. The US$1.1-billion project aims to explore basic questions such as what Jupiter is made of and whether it has a core. See go.nature.com/29icaol for more.

POLICY

EU trawling ban

A deal has been struck to limit deep-sea trawling in European Union waters, ending a long-running battle between non-governmental organizations, researchers, politicians and the fishing industry. On 30 June, ministers and Members of the European Parliament announced that all trawling below 800 metres will be banned, and fishing for deep-sea species above this depth will be permitted only in areas where it took place between 2009 and 2011. Conservation groups such as BLOOM in Paris and some scientists have long argued that deep-sea trawling destroys sensitive marine ecosystems that may take years to recover, if they recover at all.

Pesticide ruling

The European Commission has extended authorization of the use of the weedkiller glyphosate in the European Union for up to 18 months. The 29 June decision, which is pending completion of a risk assessment by the European Chemicals Agency, came a day before the previous authorization was due to expire. It followed the failure of EU member states to agree, with the necessary qualified majority, on whether glyphosate should be approved for a further 15 years — or banned. Critics fear that the chemical causes cancer, but many experts say that it is safe.

Clean-energy plan

North America is to get half of its electricity supply from non-fossil fuels and renewable energy sources by 2025, the leaders of Canada, the United States and Mexico pledged at a meeting last week in Ottawa. In their 29 June statement on a North American Climate, Clean Energy, and Environment Partnership, Canadian Prime Minister Justin Trudeau, US President Barack Obama and Mexican President Enrique Peña Nieto also agreed to reduce methane emissions, improve energy efficiency and advance carbon capture and storage technology in their countries.

Animal use ends

Medical schools in the United States and Canada have ceased using live animals to teach students. The University of Tennessee College of Medicine in Chattanooga was the last US medical school that still used animals for this purpose, and on 26 June, the university told the Physicians Committee for Responsible Medicine (PCRM) that it was ending the practice. According to the PCRM, medical schools believe that simulators provide better education as they are based on the human body.

PEOPLE

Harassment case

Michael Katze, a virologist known for his research on Ebola at the University of Washington in Seattle, has been suspended from his lab following violations of sexual-harassment policy. In a 29 June statement, the university said that it is discussing disciplinary measures. The university's two investigations found that Katze routinely made inappropriate comments to female employees and had a quid-pro-quo sexual relationship with a lab worker. The university also determined that Katze misused public resources by instructing his employees to perform personal chores for him, including soliciting sexual partners for him online.

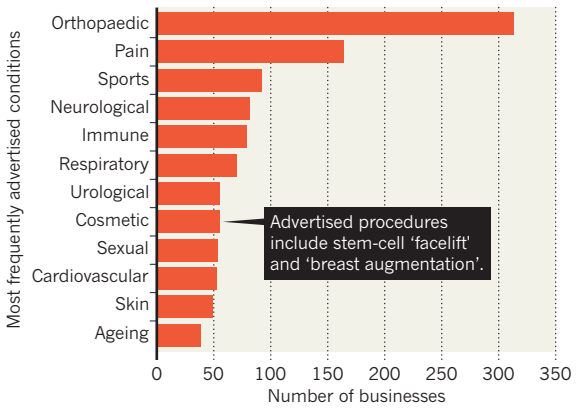
NATURE.COM
For daily news updates see:
www.nature.com/news

TREND WATCH

Hundreds of clinics in the United States are offering unapproved stem-cell treatments for a range of medical conditions. A rigorous search and analysis of Internet advertising, published on 30 June, found that 351 US businesses were marketing such treatments at 570 clinics for conditions including pain and autism (L. Turner and P. Knoepfler *Cell Stem Cell* http://doi.org/bkpv; 2016). Most clinics offered treatments with stem cells derived from the patient; other sources for the cells included amniotic and placental material.

STEM-CELL THERAPIES BRANCH OUT

Unapproved stem-cell treatments are being marketed in the United States for medical conditions as well as for cosmetic enhancement.



SOURCE: L. TURNER & P. KNOEPFLER *CELL STEM CELL* HTTP://DOI.ORG/BKPV (2016)

© 2016 Macmillan Publishers Limited. All rights reserved.

NEWS IN FOCUS

HEALTH Zika spotlights more common birth-defect virus **p.17**

CLIMATE Amazon rainforest primed for record fire season **p.18**

SPACE Small, cheap 'CubeSats' await inter-planetary ride **p.19**

EDUCATION Is it time to modernize the PhD thesis? **p.26**

GREG KENDALL-BALL/*NATURE*



John Holdren is the longest-serving presidential science adviser in US history.

POLITICS

John Holdren looks back

Obama's top scientist talks to Nature *about shrinking federal budgets, Donald Trump, and his biggest regret after nearly eight years in the White House.*

BY JEFF TOLLEFSON AND SARA REARDON

John Holdren is no stranger to the spotlight. Over his long career in science, Holdren — a physicist by training — has worked on controversial high-profile issues such as climate change and nuclear non-proliferation.

But for nearly eight years, he has enjoyed an even higher profile, as US President Barack Obama's science adviser, and director of the White House Office of Science and Technology Policy (OSTP).

With Obama due to leave the White House in January 2017, Holdren, now the longest-serving US science adviser, recently sat down with *Nature* for a wide-ranging chat. The interview has been edited for length and clarity.

Opinion polls continue to show a divide between what the American public thinks about science and what scientists think. Has Obama done enough to change the way that science is perceived?

The president has done an incredible job in making science cool for young people. This is already evident in all kinds of numbers: you see more kids enrolling in science courses, more kids participating in science fairs, more kids going to 'makerspaces'. We have substantially increased the number of engineers graduating from college in this country. I say 'we', but obviously, that is a large cooperative operation that includes colleges and universities.

I'm not sure which polls you are referring to, but my impression is that the public is more interested in and enthusiastic about science, technology and innovation than it was at the beginning of this administration.

Leaders at the National Institutes of Health (NIH) and other government agencies have

© 2016 Macmillan Publishers Limited. All rights reserved.

discussed the widespread perception that we are training too many PhDs. Do you worry about that?

If every PhD we train believes that her or his only acceptable career trajectory is a tenured professorship in a college or university, then it's true: we are training more PhDs than there are slots of that kind. But the PhD is, in fact, a very versatile degree. Far more than just demonstrating that you know more than practically anybody else about one narrow topic, it demonstrates that you have the fortitude, the focus, the commitment and the intellectual capacity to tackle a very tough problem.

PhDs are finding constructive and rewarding employment all across the economy, and, overall, our view is that there are still more opportunities for highly trained people in science, technology and innovation than there are people being trained.



BRENDAN SMIALOWSKI/AFP/GETTY

Holdren and Obama have pushed for bigger science budgets — with mixed results.

Do you worry about future science funding?

The president has consistently recommended more money for science and technology than Congress has been willing to pass.

The success ratio of proposals to the NIH is something like 17% — that is, we are funding one-sixth of the proposals that the NIH gets. And those proposals are already self-selected. Investigators don't bother writing a proposal to the NIH unless they think they have got a really good idea, a capable team and a plausible strategy. If you ask Francis Collins, the NIH director, what fraction of the proposals they get that are worthy of funding, he'll tell you 50%.

That means we are funding about a third of the potentially productive, influential, path-breaking research that is proposed to the NIH. But the NIH has a budget of over US$30 billion per year. It's not very easy in these budget times to increase a $30-billion budget by a large factor, like 50% — never mind 100% or more, as director Collins would say is warranted in terms of the quality of the research. The same is true at the National Science Foundation — far more worthy proposals than they are able to fund. This is a consistent problem. I would like to see more public support for raising public spending on research and development.

Science is global today. How do you think that complicates matters? Can the regulators keep up?

I'm going to China this week for a strategic and economic dialogue and for a US–China dialogue on innovation policy. I'll be talking with my Chinese counterpart, Wan Gang, the minister of science and technology, about some of these very problems and what we are doing about them.

We have a lot of cooperation with China on biomedical issues. We talk to them all of the time about gain-of-function research and about gene-editing issues. And in fact, when the current round of interest in gene editing emerged with the rise of the CRISPR technology, the [US] National Academies of Sciences, Engineering, and Medicine gathered leading scientists from all over the world in a format very much like Asilomar [a landmark conference in 1975 that set rules for research on recombinant DNA], but strongly international. The top Chinese people came to talk through what the implications of these technologies are, and how we should think as a global science community about regulating them.



PETE SOUZA

NATURE.COM
For an extended version of this interview, see:
go.nature.com/29mxyuj

Shortly after he took office, Obama said that this was going to be the most transparent administration ever. But journalists have found some agencies to be fairly opaque.

In the first months of the administration, the president issued executive orders on transparency, on scientific integrity, on openness in government. I was put in charge of a number of the implementation [efforts]. That has been a focus of OSTP throughout this administration. We've gotten virtually all of the departments and agencies to produce for public review and comment, and then to finalize, policies on openness and on scientific integrity. I think we've made great progress in terms of open data, in terms of the publication in open venues of federally funded research. But I would not argue that that job is finished.

There is always a tendency in government, some of it quite legitimate, not to expose internal deliberations prematurely. You know, it's quite challenging to have a discussion between the president's senior advisers with reporters from *Nature*, *Science* and *The New York Times* sitting around in the room, because if you do that, nobody will float a trial balloon for fear that the trial balloon will get into the news as a done deal.

You've spent almost eight years inside what is arguably the most powerful institution on Earth. Do you come away more or less optimistic about humanity's ability to deal with its problems?

I come away more optimistic, and that's in large measure due to the extraordinary leadership that President Obama has provided. I have felt for many decades that science, technology and innovation are crucial if human society is to get its arms around the biggest challenges we face. And I've had the pleasure of working for a president for nearly eight years now who shares that view. ■

MORE ONLINE

TOP NEWS



NASA/JPL

NASA's Juno probe enters Jupiter's orbit go.nature.com/29lmrmb

MORE NEWS

- Biomedical researchers lax about validating antibodies for experiments go.nature.com/29lmswx
- Antarctic ozone hole is on the mend go.nature.com/29ewicl
- Background does not impact UK scientists' pay go.nature.com/29jpnji

NATURE PODCAST

Landscapes and nature; the Hitomi observatory's swan song; and reforming peer review nature.com/nature/podcast

© 2016 Macmillan Publishers Limited. All rights reserved.



LOU COLLECTION/ALAMY

This 22-year-old is severely disabled by fetal cytomegalovirus infection and cannot communicate verbally.

PUBLIC HEALTH

Zika raises wider birth-defect issue

Cytomegalovirus is a greater global problem than Zika.

BY DECLAN BUTLER

A virus is killing hundreds of babies in the United States each year, and leaving thousands with debilitating birth defects, including abnormally small heads and brains. This is not the Zika virus. It is a common and much less exotic one: cytomegalovirus (CMV).

Now, as the eyes of the media and health officials focus on the spread of Zika in the Americas and beyond, many researchers and advocates hope that funders and health agencies will at last pay more attention to a much greater global problem — the millions of babies born year in, year out, with often-serious birth defects.

"Birth defects are not high on the public-health agenda," says Stanley Plotkin, a retired scientist who in the 1970s developed the current vaccine against rubella (German measles). A 1960s rubella pandemic caused tens of thousands of birth defects in the United States alone.

"Zika is an opportunity," he says — to raise the profile of birth defects among research funders and public-health agencies, and to accelerate efforts to develop a CMV vaccine.

The World Health Organization (WHO) estimates that, annually, more than a quarter of a million babies worldwide die shortly after birth from congenital anomalies, and many more are born with serious defects. The causes are many — some known, some not. A global focus on reducing child mortality has meant that severe disabilities in children are a lower public-health priority, says Anita Kar, a specialist in congenital abnormalities at the University of Pune in India.

CMV is a poster child for the problem — and with Zika so much in the news, scientists and advocacy groups are voicing frustration and trying to seize the moment. The US National CMV Foundation is running information campaigns comparing and contrasting Zika with CMV. It is lobbying politicians to build on the mandates enacted in several states for public-health authorities to produce outreach material, including billboards on sides of buses, and to do CMV tests for all infants with hearing difficulties.

"Zika has become a way to open up conversations about CMV," says Janelle Greenlee, a co-founder of the CMV foundation, who lost one daughter to congenital CMV and has another, the daughter's twin, with serious hearing loss and cerebral palsy.

CMV infections in adults, children and infants are mostly asymptomatic and harmless, but the virus is much more dangerous — often lethal — to the fetus. Worldwide, around 1 in 100 to 500 babies are born with congenital CMV, and of the 10–20% who show symptoms, about 30% will die. Survivors often have liver, lung or spleen damage, or neurological problems including developmental disability or loss of hearing or sight.

CMV's link to birth defects has been known since the 1950s — yet a 2012 survey found that only 13% of US women and 7% of men had heard of congenital CMV (M. J. Cannon *et al. Prev. Med.* **54,** 351–357; 2012). Low awareness is deadly, says Gail Harrison, an infectious-disease researcher in CMV at the Baylor College of Medicine and the Texas Children's Hospital, both in Houston.

There is no vaccine, so precautions — handwashing and avoiding contact with children's saliva and urine — are the only defence. Harrison works closely with patient groups to promote awareness, but says that she struggles with the inertia of state and federal agencies in helping to get these messages across.

The administration of US President Barack Obama has requested more than US$1 billion for research and control measures for Zika, and the website of the US Centers for Disease Control and Prevention (CDC) is awash with information and advice on that virus, she notes. But the more modest amount of information on CMV has to be actively searched for.

Leading health experts and the CDC expect that Zika in the United States will be limited to small, localized outbreaks in southern states where *Aedes aegypti*, the mosquito that transmits the virus, is present during warm parts of the year. That prediction is based on the pattern of past US outbreaks of dengue and chikungunya, two other diseases carried by the same mosquito. For the United States, says Plotkin, "there is little doubt that CMV is a bigger problem than Zika".

Contributors to birth defects include genetic abnormalities as well as many more preventable factors, such as infectious diseases, medications, diet and environmental chemicals. But the causes of almost three-quarters are unknown.

Better training in birth-defects epidemiology is urgently needed, in particular in developing countries, says Kar. Such research is difficult, requiring population-scale surveillance registries, and often relies on questionnaires that ask mothers of children with congenital abnormalities to try to recall past exposures — a process susceptible to inaccuracies.

To improve matters, pan-European and US birth-defect registries are increasingly trying to match pregnancy outcomes with vast databases of histories of prescribed drugs, local water- and air-pollution levels, and other factors. Prescription histories are especially important because pregnant women are usually excluded from clinical trials, and so little may be known about the safety of medicines for fetuses.

But many poorer countries lack even basic surveillance. In the case of neural-tube defects such as spina bifida, for example, a global review published in April found that 120 of the WHO's 194 member states had no prevalence data. "Registries are urgently required," says Kar. ■

© 2016 Macmillan Publishers Limited. All rights reserved.

POLICY

Thumbs down for 'Common Rule' revisions

Panel nixes US government changes to research ethics.

BY SARA REARDON

The US government's proposed overhaul of regulations that govern research with human participants is flawed and should be withdrawn, according to a review by the US National Academies of Sciences, Engineering, and Medicine.

The regulations — known collectively as the Common Rule — address ethical issues such as informed consent and storage of study participants' biological specimens. In their 29 June report, the academies said that the government's suggested changes are "marred by omissions and a lack of clarity" and would slow research while doing little to improve protections for patients (see go.nature.com/29afkwd). Instead, the panel recommends that an independent commission craft new rules for such research.

"This is a total smack-down," says Ellen Wright Clayton, a bioethicist and lawyer at Vanderbilt University in Nashville, Tennessee, of the academies' report.

The Common Rule, which was introduced in 1991, seeks to ensure that research with humans is ethical by minimizing patient harm and maximizing the benefit to society. Over time, achieving these goals has become more complex because of technological advances such as the rise of DNA identification, which can make it harder to maintain patient privacy.

The reforms, proposed in September by the US Department of Health and Human Services (HHS), attempt to address such emerging concerns. For instance, the HHS proposal would require participants' consent to use stored samples, such as blood or tissue, for future research. Even if samples are anonymized, the HHS says, it is fairly simple to re-identify people on the basis of their DNA.

But the US academies' panel says that the proposed consent requirements would slow research unnecessarily, because little harm is likely to come to a person as a result of the use of stored biospecimens. And if the specimens are de-identified, the extra consent forms themselves would further link the specimens to the person's name and therefore increase the risk that the person would be identified.

An HHS spokesperson says that the government is still mulling the new report and more than 2,000 public comments on its reforms. ■



Most fires in the Amazon are started by landowners trying to clear fields and forests for cultivation.

ECOLOGY

Amazon set for record fire season

Warm oceans presage intense blazes in rainforest.

BY JEFF TOLLEFSON

The Amazon is ready to burn. After an unusually dry rainy season, the southern section of the rainforest is heading into winter with the largest moisture deficit since 1998. This has set the stage for an unusually intense fire season, according to a forecast issued on 29 June that is based on sea-surface temperature trends in the Atlantic and Pacific oceans.

"The region is primed to have record fire activity," says forecast co-author Douglas Morton, a remote-sensing expert at NASA's Goddard Space Flight Center in Greenbelt, Maryland. More broadly, a team led by Morton and James Randerson, a biologist at the University of California, Irvine, says that it can predict fire risk across much of the globe — based in part on the influence of the weather pattern El Niño and its counterpart, La Niña.

The Amazon burn predictions stem from the epic El Niño weather event that emerged last year. El Niños warm the tropical Pacific Ocean, which tends to reduce rainfall during the rainy season, and the warmer temperatures in the tropical Atlantic Ocean can suppress rains during the dry season.

The El Niño that emerged last year also helped to spawn devastating forest fires in Indonesia, the researchers say. Their work reveals that sea-surface temperatures in the Atlantic and Indian oceans foreshadow fire trends in Central America, Africa and some boreal forests in Earth's high northern latitudes.

In each case, Morton and Randerson say, ocean conditions can provide a hint of precipitation trends in key forested areas on land several months in advance. "All of these processes are contributing to both the build-up of fuels and the moisture level of those fuels going into the dry season," Randerson says. "That's what leads to a predictability in global fire regimes."

FORECASTING VULNERABILITY

Other teams are looking to include fire risk in short-term and seasonal weather forecasts by incorporating independent fire models. These models attempt to account for factors such as vegetation type and the likelihood of lightning strikes or agricultural fires. Eventually, such forecasting systems could integrate more complex phenomenon such as the dynamics of vegetation growth, the way that fire tends to propagate across a landscape and the gases and particles that are emitted during a fire, says Allan Spessa, a

© 2016 Macmillan Publishers Limited. All rights reserved.

MARIO TAMA/GETTY

fire modeller at the Open University in Milton Keynes, UK.

The European Centre for Medium-Range Weather Forecasts in Reading, UK, plans to soon make public its prototype system to forecast fire risk about six weeks in advance, and the centre's modellers are working to include fire risk in their seasonal forecasts. Florian Pappenberger, who heads the centre's work on extreme-weather forecasting, says that the statistical approach used by Morton and Randerson is solid and can serve as an independent check on model forecasts, which come with their own uncertainties. Forecasts for water availability in rivers, reservoirs and agricultural systems operate in such a manner today.

"I don't think one method replaces the other," he says. "I expect that merging both will be quite beneficial."

However, whether forests actually go up in smoke depends on a host of factors, including law-enforcement and fire-suppression efforts that vary from region to region. For instance, almost all fires in the Amazon are started by landowners clearing fields and forests for cultivation and livestock. But once the humidity drops and the vegetation dries out, those agricultural fires can run wild.

READY TO BURN

The likelihood that this will happen increases as the dry season wears on, but scientists can already see El Niño's impacts. Morton and Randerson's team analysed rainfall measurements from gauges and satellites during the rainy season, and used data from NASA's Gravity Recovery and Climate Experiment (GRACE) satellites to provide an estimate of the cumulative water storage on land — in soils, aquifers and rivers — going into the dry season. Randerson says that the situation in the Amazon is worse than it was during the major droughts of 2005 and 2010 and on par with 1998, after the last major El Niño.

"All of these processes are contributing to both the building up of fuels and the moisture level."

As well as forecasting risk in the Amazon, Morton and Randerson are tracking and mapping fires there using infrared measurements collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors aboard NASA's Terra satellite. The device has detected almost 12,500 fires in the Mato Grosso region of Brazil this year alone — making 2016 the third-worst year in the MODIS record, which stretches back to 2003.

In the Amazon, the question now is whether Atlantic storm systems will bring much-needed relief during the dry season. Morton and Randerson have identified a link between Atlantic hurricanes and Amazon fires: when the tropical Atlantic is warm, cyclones are more likely to form, and those cyclones pull the rain bands that often flow into the Amazon northwards. The US National Oceanic and Atmospheric Administration's hurricane forecast currently calls for a neutral season, but the tropical Atlantic has been cooling, which bodes well for the Amazon.

"If there were to be a shift in north Atlantic sea-surface temperatures, that could short circuit this fire forecast," Morton says. ■

SPACE

CubeSats queue up for deep-space rides

Tiny craft face a wait to be propelled beyond Earth's orbit.

BY ELIZABETH GIBNEY

CubeSats — spacecraft built from 10-centimetre-sided cubes, often with off-the-shelf parts — are already ubiquitous in near-Earth orbit, doing everything from Earth observation to studies of bacterial proteins in space. Now scientists are itching to send them farther afield, and more than a dozen deep-space CubeSats are in the pipeline.

The cost — typically no more than US$10 million for an interplanetary mission — means that the mini-craft can take risks that a more costly venture could not. They can also work in swarms, which allows new kinds of experiments. CubeSats generally piggyback on the launch of other missions, and whereas trips to low-Earth orbit, such as the cargo ships that shuttle to the International Space Station, are relatively common, missions to other parts of the Solar System are much rarer.

Lifts are so hard to come by that the first interplanetary CubeSat — NASA's twin INSPIRE mini-spacecraft, intended to test key technology for future missions — has been waiting for almost two years. "We still have to find a ride," says Anthony Freeman, who manages the Innovation Foundry at NASA's Jet Propulsion Laboratory in Pasadena, California.

CubeSats were originally conceived as a teaching tool in 1999. Today, they carry out both commercial missions and near-space science. But deep space poses a much bigger challenge (see 'Miniature explorers'). Their diminutive size cannot accommodate standard propulsion and long-range communications equipment, let alone complex scientific instruments.

Engineers are starting to overcome these problems, says Roger Walker, who oversees CubeSat development at the European Space Agency (ESA). To solve the communications problem, ESA's first interplanetary CubeSats will talk to Earth through a mothership. CubeSats will take part in the joint ESA–NASA Asteroid Impact and Deflection Assessment (AIDA) mission, planned for 2020, where they will take on risky jobs such as up-close data collection as a larger probe plunges into an asteroid.

NASA's planned mission to Europa, currently under development, would also use the mother–daughter model, deploying a fleet of CubeSats to make close fly-bys of the Jovian moon. Scientists think that Europa could harbour life under its icy surface.

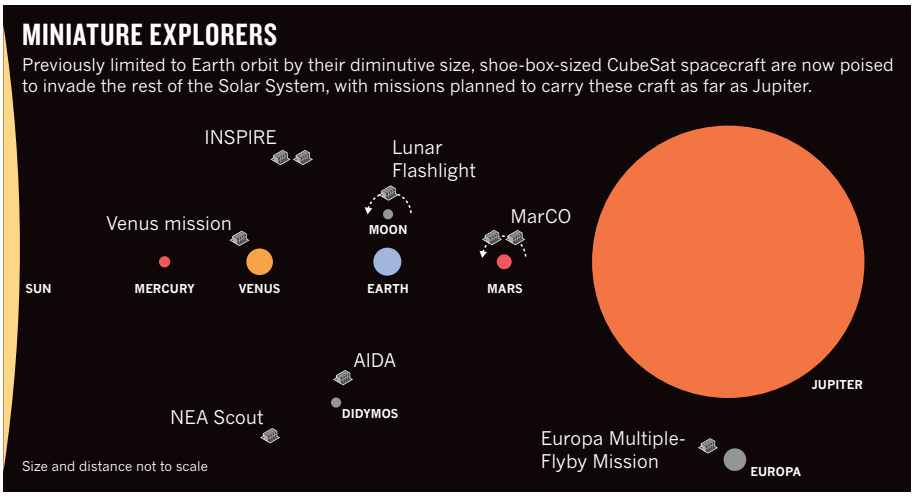
Lone deep-space CubeSat missions are also on the horizon. NASA has developed a miniature radio-communication system capable of talking directly to Earth from Mars and beyond. The agency will test the system on INSPIRE — which has a side-mission of mapping interactions between Earth's magnetic field and the solar wind — and on Mars Cube One (MarCO), twin communication satellites scheduled to fly on the InSight mission to Mars when it launches in 2018 after a two-year delay. NASA has also developed tiny, cold-gas firing thrusters for propulsion, and radiation-resistant electronics that can survive beyond the protection of Earth's magnetic field.

Meanwhile, firms in Europe are developing high-efficiency ion engines, and a company in Rome called IMT is looking at ways to power such engines with deployable solar panels that can turn to constantly face the Sun. Together, all these technologies make solo CubeSats missions feasible, says Walker.

Freeman predicts that more than a hundred CubeSats could be dispatched throughout the Solar System by the end of the next decade — but only if they can get into space. He is calling on all space agencies to agree to carry at least one CubeSat on each major planetary mission. Walker agrees: "It would really stimulate the area. Ultimately, that's the main problem to overcome for interplanetary CubeSats, alongside communications." This would mean forging plans for a CubeSat tag-along early in the mission's design phase.

To cope with the large number of CubeSat proposals, NASA also wants to see more ▶

© 2016 Macmillan Publishers Limited. All rights reserved.



low-cost commercial launchers developed, to carry tens to hundreds of kilograms of payload, in contrast to the 5 tonnes typical of launchers designed for communications satellites. Freeman says that such smaller rockets could carry perhaps a few dozen 5-kg CubeSats to low-Earth orbit, or be adapted to include an upper stage that could take a single CubeSat to deep space. He hopes to use a similar method to send a free-flying probe to Venus, where it would skim through the planet's acidic atmosphere.

CubeSats aimed for the Moon might get an easier ride. NASA's Space Launch System, a heavy-lift rocket designed to send people beyond Earth's orbit, will carry 13 CubeSats and an uncrewed Orion capsule on its maiden launch in 2018. The cargo will include Lunar Flashlight, which will use a reflected beam of light to look for icy deposits in the Moon's dark craters, and Near-Earth Asteroid (NEA) Scout, designed to explore a nearby asteroid.

ESA is developing a separate lunar approach. Together with Surrey Satellite Technology Ltd (SSTL) in Guildford, UK, and the Goonhilly Earth Station in Helston, UK, it is developing a system that could solve two problems: a commercial mothership that would provide transport to the Moon and a data relay for dozens of CubeSats, for a fee of around £5 million (US$6.6 million) per craft. Eventually, such a model could expand, says the SSTL's Christopher Saunders. "Essentially, we want to build a Solar-System internet," he told the Interplanetary CubeSat workshop in Oxford in late May.

According to Freeman, CubeSats will soon be able to carry instruments that would have seemed off-limits only a few years ago, such as high-resolution imagers and radar altimeters. And a recent investigation by the US National Academies of Sciences, Engineering and Medicine of CubeSats' potential concluded that the probes are capable of doing "fantastic science", Thomas Zurbuchen, a space scientist at the University of Michigan in Ann Arbor, said at the meeting. "Much of it has yet to be imagined." ■

CLARIFICATION

In the News Feature 'Mystery in the heavens' (*Nature* **534,** 610–612; 2016), the discussion of the initial radio burst meant to say that over the course of just a few milliseconds, the source's output matched that of 500 million Suns in the same time period.

© 2016 Macmillan Publishers Limited. All rights reserved.

BACK TO THE THESIS

Late nights, typos, self-doubt and despair. Three leading scientists dust off their theses, and reflect on what the PhD was like for them.

BY KERRI SMITH & NOAH BAKER



CHRIS MADDALONI/*NATURE*

Francis Collins shakes his head in bewilderment as he flicks through the pages of his thesis. "At this point it looks very much like another language," he says, looking with puzzlement at page 71, which contains far more equations than text. The PhD was on theoretical quantum chemistry, and had "absolutely no practical application", Collins says. Looking at it now, "it does feel a little bit like this was another person".

Collins was in his early 20s when he was studying for his doctorate at Yale University in New Haven, Connecticut, modelling how small groups of atoms interact. "A lot of what I did was pencil on paper, trying to solve really complicated calculus equations. It was a little lonely at times," he says. Then, about halfway through his studies, he decided to quit his PhD and transfer to medical school. He ended up finishing the thesis in his spare time. "I spent many nights and many weekends trying to get this written out," he says, with something of a grimace. "I made myself a schedule and tried to stick to it, with my little electric typewriter, banging away."

The writing machines have changed, but the slog is the same. Completing a thesis is a huge undertaking for PhD students, and many struggle to get that far: only around 70% of UK students who embark on doctoral studies actually emerge with a PhD, and the rate is just 50% in the United States. Many of those who do finish move on to careers outside academia; even those who stay sometimes wish they'd spent more time writing papers — the currency of career progression — instead (see page 26).

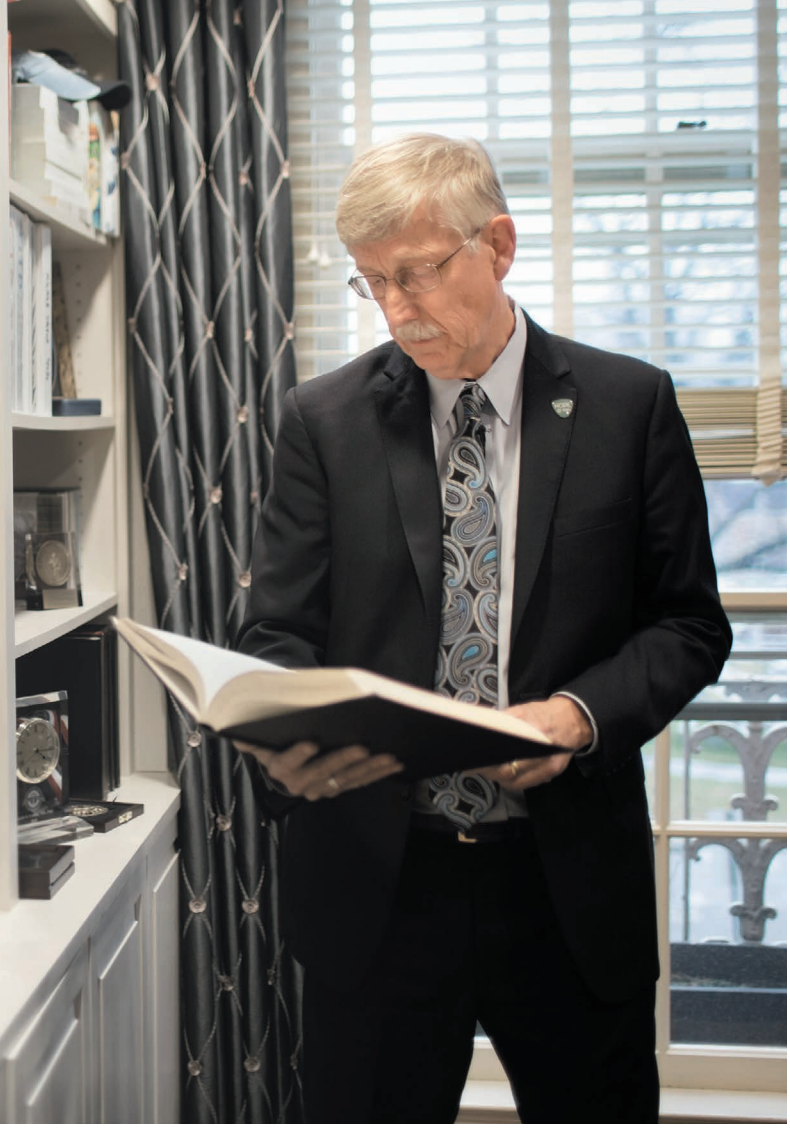
So what value does the thesis retain, and what lessons does completing one impart? To find out, *Nature* asked three prominent scientists to dig out their theses, thumb through the pages and reflect on what they — and the world — gained from them. What did they learn that could be of value to students who are writing up today? Their reflections, sometimes surprising, are recorded in three short films that accompany this article online (see go.nature.com/297qrah).

Collins's PhD was the start of a stellar career: he famously moved into biological research, identified the gene that causes cystic fibrosis, led the Human Genome Project to completion and now, more than 40 years after writing his thesis, directs the US National Institutes of Health. But that doesn't mean that his PhD changed the world. "Did it really add significantly to the knowledge the Universe contains?" he says. "Well, it would be a rather small contribution, to be sure."

But like others who went 'back to the thesis' for *Nature*, he thinks that what matters was not so much the subject or results, but what he learnt about the process of research along the way. "I think the greatest beneficiary of my PhD was not the Universe," Collins says. "It was probably me."

NATURE.COM
Watch Collins, Seager and Frith talk about their theses at go.nature.com/297qrah

© 2016 Macmillan Publishers Limited. All rights reserved.



FRANCIS COLLINS

SEMICLASSICAL THEORY OF VIBRATIONALLY INELASTIC SCATTERING, WITH APPLICATION TO H^+ AND H_2 (1974)

Collins keeps his PhD on a low shelf in his office, next to those from his past students. He pulls it out and places a paternal hand on the thick, leather-bound book. "I think it turned out pretty well. It's quite a hefty document," he says.

The road to that document started back in 1970, when Collins arrived in the lab of Jim Cross, a theoretical chemist at Yale. Cross remembers Collins as "a quiet, unassuming man, not particularly sophisticated culturally". But, he says, "I quickly realized that he was one of the brightest and most broadly based students that I have ever met".

Collins was tasked with developing theoretical models to explain what happens when a proton is fired at a hydrogen molecule: how does the energy of the two bodies dissipate, and could the hydrogen be coaxed into another state? Day after day, he sat at his basement desk, tackling calculus equations and writing corresponding computer programs in Fortran. He used a machine in the university computer centre to punch the programs onto cards, then waited until after 1 a.m., when electricity was cheaper, to feed the cards into the mainframe computer. "It did make me begin to wonder, OK, is this the right path for me?"

It wasn't — something Collins came to realize during an all-nighter about halfway through his studies. He was talking to a fellow graduate student, Jay Gralla, who was examining how molecules of RNA fold up into secondary structures. The broader aim was to understand the rules by which genetic information in RNA and DNA is used to build biological systems. Collins was blown away. "I was astounded that I had missed this whole thing about biology — that it was digital, it was an information system, it did have principles," he says. "It was a revelation."

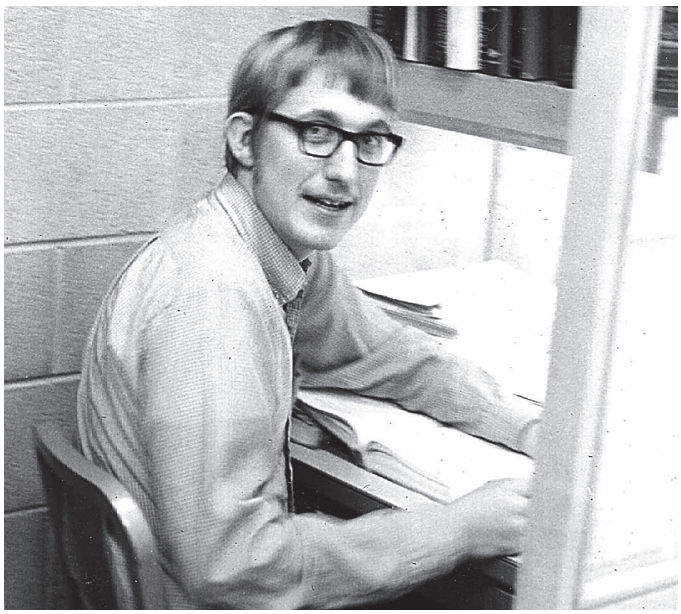
Shortly afterwards, Collins decided to switch to medical school. "That was a wrenching time," he says. He was drawn to explore biology and medicine, but he also had a growing family, financial strains and "all kinds of self-doubts". He also didn't know if he'd actually done enough work to complete his PhD — but Cross told him to write it up anyway. Collins stayed behind in New Haven to write, while his wife and young daughter left for the family's new home in North Carolina. He still couldn't get it done before his medical studies started. By the time of his graduation ceremony, in May 1974, he was finishing his first year of medical school and expecting a second child. He didn't attend.

Several years later, his medical training complete, Collins returned to Yale to work in a molecular-biology lab, and never looked back. The exactitude instilled in him by his PhD stayed with him. He had learned to assess a complex system, strip it down to its component parts and glean insights from it. "That's something that I do now in my lab," he says.

His thesis work isn't in much demand. The model of colliding particles that Collins developed was a good match with others' experimental findings, and made some useful approximations, but the work has been convincingly superseded by advances in processing power. "These days, a theoretical chemist wouldn't dream of limiting themselves by doing these approximations," he says.

Reflecting on it now, Collins is glad that he took the risk of switching fields, and would encourage today's PhD students to take chances too. Transitions in a career are "when you grow the fastest; they're when you're really alive". And think big, he urges. "If you're going to study something, study something important. It might be risky, it might be hard, it might not work, but there are too many people spending their time on obvious next steps."

"IT DOES FEEL A LITTLE BIT LIKE THIS WAS ANOTHER PERSON."



COURTESY OF KATHERINE ALBEN

© 2016 Macmillan Publishers Limited. All rights reserved.

SARA SEAGER

EXTRASOLAR GIANT PLANETS UNDER STRONG STELLAR IRRADIATION (1999)

A flicker of embarrassment crosses Sara Seager's face when she is asked whether there are any mistakes in her thesis. "I definitely have at least one typo. I know where it is, unfortunately. I hate to talk about it." She thinks for a moment, her thesis unopened on the desk before her. "Now that you mention it, I should probably go back and correct it with a pen."

There is little else for Seager to regret about her thesis. She is now a planetary scientist at the Massachusetts Institute of Technology in Cambridge, and, unusually, her PhD helped to found a field. "It might have been one of the first — if not the first — PhD theses on exoplanets," she says.

In 1996, when Seager started her postgraduate studies at Harvard University in Cambridge, just half a dozen planets had been spotted orbiting distant suns. They could be detected only indirectly, mostly by capturing the 'wobble' that an orbiting planet caused in the movement of a star. And the signals were noisy — some astronomers didn't believe that exoplanets were real.

Seager was encouraged to enter the field by her supervisor at Harvard, Dimitar Sasselov, who was keen to take a different approach. Sasselov encouraged Seager to study the atmospheres of exoplanets to find ones that might harbour interesting chemistry or indicate life. This seemed unlikely to work when the planets themselves were so difficult to detect. "It was a big risk at the time: a non-tenured professor and a grad student. Despite the advice otherwise of colleagues in the department, we went ahead," Sasselov says.

Seager built a theoretical model suggesting that it should be possible to see starlight bouncing off a planet that was orbiting its star closely, and that analysing that light would reveal a fingerprint of the planet's chemical constituents, temperature and pressure[1]. Shortly afterwards, during her postdoc, she predicted that it should be possible to spot clouds in the atmosphere, and that one of the easiest elements to detect would be sodium[2].

It was tough going. She derived equations to represent the components of a planet's atmosphere and then, after teaching herself to code, plugged them into the computer models she was building. Her hours were long and isolated, and she would often hit programming bugs that threatened to derail her work. Meanwhile, ex-students from her department were calling from Silicon Valley: their companies were seeking people like her. "I was far from committed to a career in science. I often thought of leaving," she says.

Yet Seager "always expressed a certainty about what she was working on", recalls David Charbonneau, a contemporary of hers at Harvard who now leads an astronomy group there, and was using Seager's theoretical predictions to explain his observational results. He describes her as a



NOAH BAKER/*NATURE*

UTA FRITH

PATTERN DETECTION IN NORMAL AND AUTISTIC CHILDREN (1968)

"I have not looked at this in decades," declares Uta Frith as she retrieves her thesis from a study in her Victorian house in suburban London. The book, bound in sky-blue cloth, nestles on a low shelf, right next to a science-fiction encyclopaedia. She dusts it off with a cloth and opens it to the typewritten title page. "It looks very charming and childish. That's really my immediate impression. I did want a short and an interesting title."

The title is as brief as Frith's PhD was: she had only two years of funding, starting in September 1966, and at the end of 1968 she duly turned in the thesis: 205 pages, typed up by a secretary from her handwritten manuscript. The bibliography is concise, just 10 generously spaced pages. "So little was known about autism at the time that this was the extent of the references I found," she says. Today, the developmental disorder is the subject of several thousand publications each year.

Frith came to London from her native Germany in 1964 to attend a course in abnormal psychology at the Institute of Psychiatry. There, for the first time, she met children with autism, and was "completely fascinated. I still am," she says. She also met her future supervisors, psychologists Beate Hermelin and Neil O'Connor. At that time, autism spectrum disorders were poorly understood and carried a stigma. Those diagnosed were usually only the severe cases, children with profound intellectual and linguistic difficulties. The mainstream view in psychiatry was that autism was a product of a child's upbringing and environment and that distant, unloving parents — particularly mothers — were to blame.

Frith refused to subscribe to that view. "I was always struck, when I met the parents of these children, how little they corresponded to what was told about them in the literature," she says. The question that interested Frith was whether the children might process information differently from other kids. To investigate this, she showed children a simple box containing green and yellow counters that were arranged in a specific pattern. She then covered up the box and asked the child to build the sequence from memory.

She often travelled to hospitals to test children with autism, as well as to nurseries and schools to assess children in her control group. She plugged the data into mechanical calculators that were "very, very noisy" and then took the better part of a day to perform the statistical analysis.

Frith turns to the later pages of her thesis to remind herself of what she found, well aware of how dated — even naive — it might sound. "I'm a little bit afraid of this now. What nonsense can it be?" She

© 2016 Macmillan Publishers Limited. All rights reserved.



NOAH BAKER/*NATURE*

fierce intellectual and recalls how annoying she found any imperfection. "She would get frustrated if the data weren't as unambiguous as she would have liked."

Seager says that the day she got her computer code to work "was one of the defining moments of my entire life". And once her work was finished, she didn't have to wait long for her predictions to be tested: in 2002, astronomers including Charbonneau detected the first exo-atmosphere[3], and found that it contained the sodium signature, albeit at a slightly lower level than Seager had predicted. Since then, the field has flourished: 3,285 exoplanets have now been confirmed, and the study of their atmospheres has bloomed. The material in Seager's PhD has been used by astronomers to request time on the Hubble Space Telescope, Keck observatory and other instruments. And although Seager can't erase the sole typo from her thesis, she points out that the papers she published from it are free of mistakes.

Did Seager enjoy her PhD? "Unfortunately, I think the answer might be no." But she does have fond memories of the time she spent writing up her research. "I remember when I was finishing it, I didn't go to any other talks, I didn't really read the news, it was just put the blinkers on and get the job done." She found great satisfaction in devoting herself to a single task, and relished the clarity of thinking that afforded her. "The world goes away", she says. "And so when you're in that zone, actually you're happy."

Seager now tries to make sure that those in her lab have the space to think too. "I do let the students spin their wheels. They have to, or they won't find their own way." And if she could give advice to her younger self, it would be simply: "Hang in there."

As for the thesis itself — a slim, red volume with gold lettering — it's not something she feels sentimental about. "I've met people who, they cry when they give away their kids' baby clothes, but I was never one of those — and I think I felt the same way about the thesis." She's more inclined to look forward. "In exoplanets, the best planet, the best discovery, is the next discovery."

COURTESY OF UTA FRITH

showed that children with and without autism both made errors in about 25% of the trials, but that they made different mistakes. Children in the control group tended to follow the pattern too strongly — perhaps placing three green counters together instead of two. Those with autism, however, placed the counters in their own simple pattern, such as green, yellow, green, yellow. Frith proposed that these children impose very strict patterns on the outside world, too, and this idea seemed to correlate with the behaviours that clinicians at the time considered characteristic of the condition — obsessions with particular objects, for example, or disliking change.

Frith saw logic in the children's responses, and felt that they were not necessarily inferior to those of others. "It is presumptuous to think that those patterns imposed by autistic children are any worse than the patterns I have imposed on the data," the concluding paragraph of her thesis reads. "Well, that's quite philosophical," she says, in modest delight.

Frith is aware that she was studying at a golden time. Psychology was thriving in the United Kingdom; she had the undivided attention of two supervisors; and, just as she was coming to the end of her PhD, she was offered a full-time job at a Medical Research Council (MRC) unit where one of her supervisors had just been appointed director. "I was just so fortunate," she says. The post led to a 50-year career with the MRC and University College London, during which Frith showed that children with autism have deficits in their 'theory of mind', the cognitive capacity to understand that others have their own beliefs and ideas. This was an important concept that was "just emerging in primate work" and that she adapted to studies of autism, says Ami Klin, who directs the Marcus Autism Centre at Emory University in Atlanta, Georgia, and whose 1998 PhD was co-supervised by Frith. "She was always extraordinarily open-minded, patient, supportive," he says.

Frith knows that today's PhD students have a much tougher time: funding is tight and academic jobs scarce. But she remains a fan of the PhD as an apprenticeship in research. She learned from scratch how to formulate hypotheses, design experiments and analyse data. "It does mean doing what we might call slave labour for some of the time, but you learn through that, and you can see what it feels like to be a scientist," she says.

Uta Frith and her supervisor Neil O'Connor, in 1971.



"I'M QUITE SURE THAT IT WOULD NOT MEET REQUIREMENTS NOW."

She admits that her thesis is a product of a different era — "I'm quite sure it would not meet the requirements now" — and she is willing to bet that there are mistakes in the text. "But who knows? I haven't read it. Why should I? There's so much else more interesting to read." ■

Kerri Smith *and* **Noah Baker** *are multimedia editors for* Nature *in London.*

1. Seager, S. & Sasselov, D. D. *Astrophys. J.* **502,** L157–L161 (1998).
2. Seager, S. & Sasselov, D. D. *Astrophys. J.* **537,** 916–921 (2000).
3. Charbonneau, D., Brown, T. M., Noyes, R. W. & Gilliland, R. L. *Astrophys. J.* **568,** 377–384 (2002).

© 2016 Macmillan Publishers Limited. All rights reserved.



FUTURE OF THE THESIS

PhD courses are slowly being modernized. Now the thesis and viva need to catch up.

BY JULIE GOULD

On the morning of Tom Marshall's PhD defence, he put on the suit he had bought for the occasion and climbed onto the stage in front of a 50-strong audience, including his parents and 6 examiners. He gave a 15-minute-long presentation, then faced an hour of cross-examination about his past 5 years of neuroscience research at the Donders Institute for Brain, Cognition and Behaviour in Nijmegen, the Netherlands. A lot was at stake: this oral examination would determine whether he passed or failed. "At the one-hour mark someone came in, banged a stick on the floor and said '*hora est*'," says Marshall — the ceremonial call that his time was up. "But I couldn't. I had enjoyed the whole experience far too much, and ended up talking for a few extra minutes."

Marshall's elaborate, public PhD assessment is very different from that faced by Kelsie Long, an Earth-sciences PhD candidate at the Australian National University (ANU) in Canberra. Her PhD will be assessed solely on her written thesis, which will be mailed off to examiners and returned with comments. She will do a public presentation of her work later this year, but it won't affect her final result. "It almost feels like a rite of passage," she says.

PhDs are assessed in very different ways around the world. Almost all involve a written thesis, but those come in many forms. In the United Kingdom, they are usually monographs, long explanations of a student's work; in Scandinavia, science students typically top-and-tail a series of their publications. The accompanying oral examination — also called a viva voce or defence — can be a public lecture, a private discussion or not happen at all. There is wide variation across disciplines and from one institution to the next. "It is a complicated world in doctoral education.

ILLUSTRATION BY OLIVER MUNDAY

© 2016 Macmillan Publishers Limited. All rights reserved.

One format does not fit all," says Maresi Nerad, founding director of the Center for Innovation and Research in Graduate Education at the University of Washington in Seattle.

This isn't necessarily a problem in itself, but some researchers worry that the decades-old doctoral assessment system is showing strain. Time-pressured examiners sometimes lack training and preparation for PhD assessments, which can lead to lack of rigour. "Two or three examiners come together to go through the thesis in a perfunctory way. They tick the boxes, everyone is happy, and then a PhD walks away," says Jeremy Farrar, director of the biomedical research charity the Wellcome Trust in London.

Farrar, like other scientists, suspects that the PhD assessment is not keeping up with the times. Single-author tomes seem outdated when much of research has become a multidisciplinary, team endeavour. Research is becoming more open, but PhD assessments can lack transparency: vivas are sometimes held behind closed doors. Some PhD theses languish, little-used, on office shelves or in archives. "We're seeing some students who are still submitting paper theses to us — they don't have electronic theses yet," says Austin McLean, director of scholarly communication and dissertation publishing at ProQuest in Ann Arbor, Michigan, which has the largest database of PhD theses in the world. What's more, little attention is given in the PhD assessment to soft skills such as management, entrepreneurship and teamwork, even though these are an essential part of life beyond the PhD, and students are increasingly leading that life outside academia. "The assessment of the PhD hasn't been updated to fit the modern definition of a PhD," Farrar says.

> "'The thicker my PhD, the better' has become a myth in the PhD community."

"There are a lot of pressures to make changes to the thesis," says Suzanne Ortega, president of the Council of Graduate Schools in Washington DC, one of a number of groups discussing the issue. The council organized a workshop in January this year called Future of the PhD Dissertation, and in March, the Australian Council of Learned Academies (ACOLA) in Melbourne examined changes to the thesis as part of a review on researcher training. Some scientists and education experts welcome the attention. "I don't think the current model for thesis examination is ideal, but there are positive movements towards changing it," says Inger Mewburn, director of research training at ANU and editor of the blog The Thesis Whisperer, which is dedicated to those completing a thesis.

PASSING THE TEST

Academics agree about one thing regarding the PhD assessment — its aim. The traditional goal is to demonstrate the candidate's ability to conduct independent research on a novel concept and to communicate the results in an accessible way. Where the academics differ is on how best to achieve that goal.

Shirley Tilghman, a molecular biologist and former president of Princeton University in New Jersey, sees merit in the monograph form of the thesis. It demonstrates scholarly ability by requiring students to "frame the historical context of a problem, describe in detail the purpose and execution and then come to a credible conclusion", she says.

But should the thesis include academic publications, too? That's the norm at the Karolinska Institute in Stockholm, Sweden, where most theses are a compilation of the student's original papers, along with a relatively short discussion, perhaps 50 pages long. The rationale is that publishing should be part of training because it better equips students for academic life and securing jobs.

Some students who complete a monograph end up wishing that they had spent more time on writing papers. James Lewis successfully defended his physics PhD at Imperial College London in October 2015, but he thinks that his one published paper landed him his postdoc at NASA's Goddard Space Flight Center in Greenbelt, Maryland. "The job market for postdoc positions is very competitive," he says, "so if you can get a paper published during your PhD then you're helping yourself." While he waits to start, Lewis is spending his days writing papers based on his research. "I'm wondering: would it not have been better to write these instead of the thesis, which took me five months to write?"

But others argue that the pressure to publish could rob PhD students of valuable parts of their studies, such as the time to shape their research path and to think creatively and independently (see page 22). "The PhD might become driven by papers only," says Farrar. "Students might end up spending their time focusing only on what papers they can produce, then staple them together with a summary and they're done — adding to the sense that the whole scientific enterprise is a paper factory rather than an exploration."

Long is working at ANU towards a thesis-by-publication: she's written and submitted one paper and has started on a second. But she's struggling. "I am finding this one much harder to write, mostly because it isn't as new or exciting as the previous one," she says. What's more, her strategy depends on things at least partly outside her control — on her PhD generating enough complete studies for publication and on a reasonably timely peer-review process.

Completed PhD theses are typically stored in university libraries — but that doesn't mean that they are read or used. Some 60% of submissions to the ProQuest database fall under the category of science, technology or mathematics, but they are the ones that are accessed least. "We think this is because the communication is more journal-focused," says McLean. Scientists do tend to keep a copy of their theses in their office or lab for use by students and colleagues. Neil Curson, a physicist at the London Centre for Nanotechnology, says that his PhD, written more than 20 years ago, is still consulted by his students when they come into his lab. Many theses, however, end up collecting dust.

VIVA LA VIVA

Whatever form the thesis takes, it has to be assessed — in most countries, by a panel of experts, and often involving an oral exam. But the viva "doesn't have the same level of consistency as the written form of examination", says Allyson Holbrook, an education researcher at Australia's University of Newcastle. In Israel, the viva is optional and very few students choose to have one; in the Netherlands, it is formal and ceremonial; in the United Kingdom, it's typically a private affair with two or three examiners; and in Australia, it's hardly performed at all. "One hundred per cent of the doctoral examination is about the thesis here," says Holbrook. That's largely because, historically, there weren't enough experts in the country to examine the work in person and it was costly to fly them in, she says.

Holbrook and her research team published a study last year that compared the assessment methods used in Australia with those in New Zealand and the United Kingdom (T. Lovat *et al. Higher Ed. Rev.* **47,** 5–23; 2015). They concluded that doing an oral defence rarely changed the result, and that the thesis itself was the "determinative step" of passing. The review on Australian research training published in March didn't support adding a viva either, but it did recommend a move towards more continuous assessment of a student, rather than waiting until the end of the training.

Some researchers see problems with the viva. It's not uncommon for nerves to get the better of a student, and for them to freeze in front of their audience, however small it is. Examiners could worsen the situation by

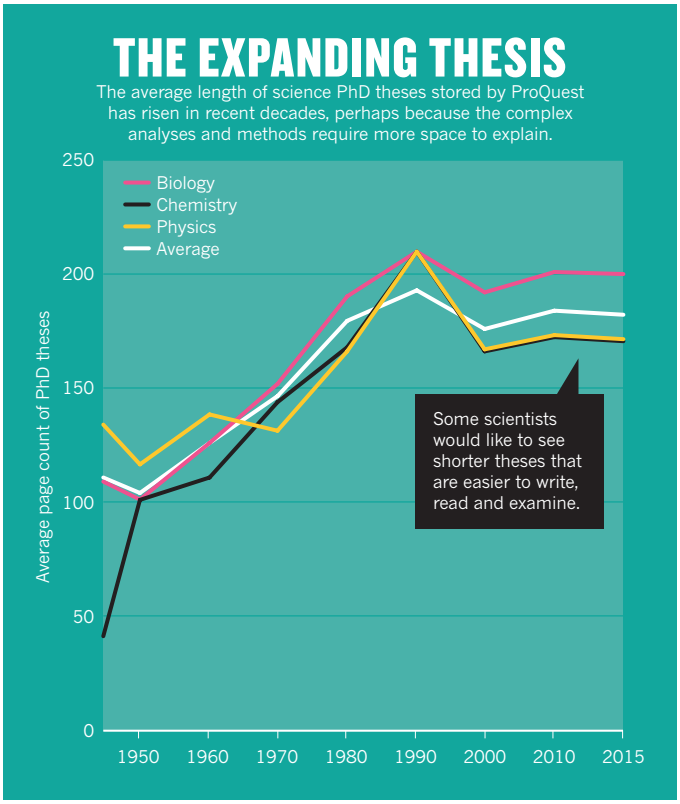
© 2016 Macmillan Publishers Limited. All rights reserved.

asking very difficult questions, says David Bogle, a chemical engineer at University College London. "There are cases where undue pressure is placed on the candidate by the examiners. This shouldn't be allowed."

TRIAL BY ERROR

Most researchers don't support a global standard for the PhD assessment. A one-size-fits-all approach would be impossible to implement, they say, and the type of assessment — be it continuous appraisal, written thesis or oral exam — should depend on discipline, project, student, supervisor and institution. "If you take away the variability in assessment and form of the thesis then you lose all creativity and innovation from the PhD," Nerad says.

But many feel that the system could be improved — by making the thesis shorter, for one. Data from ProQuest, which stores 4 million theses, show that the average length of biology, chemistry and physics PhDs soared to nearly 200 pages between 1945 and 1990. That could be because students are analysing more complex questions, performing longer literature reviews and using increasingly complicated methods that require lengthier explanations (see 'The expanding thesis'). "It's unnecessary to have such a long thesis," says Farrar, who recently assessed one such tome. "'The thicker my PhD, the better' has become a myth in the PhD community, and is taking it down the wrong direction."



SOURCE: PROQUEST

Farrar says that a slimmed-down document would be more appropriate. That could follow the concise format of a research paper, and include a review of the field, then short chapters on methods, analysis and discussion. "It would be more succinct and focused. And the examiners will probably read it all."

That isn't necessarily the case now. Examiners have to find time to review theses in between research, teaching, grant-writing and many other demands. "Something has to give, and what gives is the amount of time spent on any of those individual tasks," says Farrar. That means an examiner might skim through years of a PhD student's work in just a couple of hours. "I think we owe it to the students to examine them properly and help prepare them for their future careers," he says.

THE MODERN THESIS

One way to better reflect the team-based nature of science would be to write a joint thesis, an approach that has been used in arts and humanities graduate education in the past. However, this can make it difficult to assign credit. "If you have worked on a collaborative dissertation, a potential employer might struggle to see whether you really are an independent thinker or could you read a lead a research project," says Ortega.

There is another matter to wrestle with — the fact that half of science PhD graduates in the United States are choosing careers outside of academia, according to the National Science Foundation's 2014 Survey of Earned Doctorates. "Under those conditions, the standard assessment should include the skills in what they'll need when going on to future careers," says Michael Teitelbaum, labour economist at Harvard Law School.

Increasingly, institutions offer courses to PhD students in skills such as teamwork, management and research ethics, but these skills aren't usually assessed formally. The viva would be one opportunity to do so, perhaps by seeing how students react to various scenarios. Alternatively, as the ACOLA review suggested, PhD candidates could accrue credits in transferable skills through professional-development activities that are recorded in a portfolio. "You can't just assume that if you throw them into an environment they will meaningfully learn from that environment," says psychologist Michael Mumford, a director of the Center for Applied Social Research at the University of Oklahoma in Norman. "We need exams that ask students to deal with both real-world problems as well as ambiguous academic problems."

Farrar thinks that a change in emphasis could help. Rather than thinking of the thesis and viva as an exam, it should be viewed as the culmination of a long project. "You need to look at the PhD in the context of those four years of research, not just as revision for one big test."

Mewburn stresses that whatever form the assessment takes, it should focus more on the individual than on their work. "My preference is to assess the researcher," she says, "but we haven't developed the tools and curriculum to do that."

FEW FAILURES

It's difficult to find figures on how many students fail their PhD if they get to the point of submitting a thesis but, anecdotally, scientists say that few flunk it outright. More often, students are sent away with minor or major corrections that have to be completed before the PhD is awarded.

There are theories that few students fail because universities want to keep their number of graduates high for the rankings. But most researchers dispute this, and point to other reasons. One is that weak students are likely to have dropped out before they reach the final assessment. Furthermore, supervisors and the supporting institutions typically work hard — through regular reviews and assessments — to make sure that a candidate and project are of a sufficient standard before the thesis is submitted. "You haven't done your due diligence as a university if a student is getting to a stage where they are sending out theses that are going to fail," says Simon Hay, a global-health researcher at the University of Washington.

Nerad sees no need to reform the final PhD assessment. For her, the problem lies with the variability of graduate education as a whole. "Now that research is becoming more globalized, the PhD needs to be too." That process is under way, Nerad says: the pressures of economic globalization, international policies and national drives to house world-class universities have led to a more standardized PhD experience across the world.

During her tenure as Princeton's president, Tilghman was often asked if there was a perfect way to assess a PhD course. Not many liked her answer — that she could only really evaluate a student at the 25-year reunion. "In the end, the only way you can assess it is whether the graduates of the programme become successful scientists. If they do, you've done a good job. If they haven't, you haven't." ■

Julie Gould *is an editor for* Naturejobs.

© 2016 Macmillan Publishers Limited. All rights reserved.

COMMENT

PUBLISHING Lessons from 30 years of trying to improve peer review **p.31**

LANDSCAPE From 'Capability' Brown to the US National Park Service **p.34**

POLICY Fetal-tissue research is key to infant health, say subpoenaed researchers **p.37**

SOCIAL SCIENCE Physicists, biologists and chemists, why so snooty? **p.37**



KEVIN FRAYER/GETTY

The greater availability of data on air quality has gripped the public, especially in heavily polluted cities such as Beijing.

Validate personal air-pollution sensors

Alastair Lewis and **Peter Edwards** call on researchers to test the accuracy of low-cost monitoring devices before regulators are flooded with questionable air-quality data.

The public is increasingly aware of the health and economic costs of air pollution. Poor air quality is linked to over three million deaths each year, and 96% of people in large cities are exposed to pollutant levels that are above recommended limits[1]. The costs of urban air pollution amount to 2% of gross domestic product in developed countries and 5% in developing countries (see go.nature.com/28qv0ka).

Media attention and the increasing availability of data are reinvigorating efforts in many countries to tackle air pollution, driven as much by local and national politics as by science.

In response, start-up companies are rushing to produce cheap air-monitoring sensors, costing hundreds rather than tens of thousands of pounds. Such devices bridge gaps between sparse government measurements and individuals' wishes to track their personal exposures[2]. In a wealthy city, a single official monitoring station might represent 100,000 people; in emerging economies, one instrument covers millions of citizens.

Although personal sensors have not yet achieved their market potential, applications are promising. Portable sensors are becoming a mainstay of health research by showing people's exposure to environmental factors ranging from noise to particulate matter[3,4]. Live pollution data can be integrated into traffic-management systems to track the impacts of policies such as low-emissions zones. Affordable air-quality devices are being produced for developing countries. For example, the ▶

© 2016 Macmillan Publishers Limited. All rights reserved.

United Nations Environment Programme launched a device in 2015 at a modest cost (around US$1,500) to measure particulates, sulfur and nitrogen oxides as part of a government pilot scheme in Kenya.

All this excitement presumes that these low-cost air-pollution sensors are fit for purpose. For regulatory applications, governments and scientists use the most accurate, but expensive, detectors. And although the interpretation of the data is a subject of lively debate, the quality of readings is rarely questioned. By contrast, few of these low-cost devices have been rigorously tested and most researchers view the buzz as being beyond the serious business of academia.

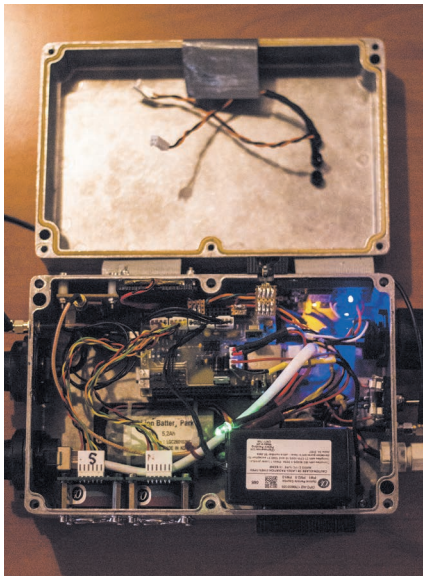
The research and regulatory communities are behind the curve. The penetration of these devices into the public domain, generating large volumes of untested and questionable data available to all, is inevitable and will increasingly become a headache for those who are responsible for managing air quality. And opportunities beckon. Atmospheric chemists must engage so that these technologies can realize their huge potential.

COMPLEX BLEND

Measuring atmospheric pollutants is challenging. Most gaseous pollutants, such as nitrogen dioxide (NO_2) or ozone, occur at parts-per-billion levels in air and are blended with thousands of other compounds. Unburnt fuel, for example, contributes many different hydrocarbons to the urban atmospheric mix. Added to this are large and changeable amounts of water vapour and carbon dioxide, at temperatures anywhere between –30 °C and 50 °C. This is difficult analytical chemistry at the best of times.

Atmospheric chemistry research has long been a hotbed of invention for detection technologies and analysis methods. Ideas emerged mainly from universities, institutes and a few research-led companies, such as Aerodyne Research and Picarro in the United States and Ionicon in Austria. The fruits of this labour have been tested by peer review; there are entire journals devoted to atmospheric instruments. Fresh technologies must establish credentials. The best ones are absorbed by a few early-adopter research groups. Over perhaps a decade, successful methods find their way into research use; a rare few make it into regulatory networks. Along the way come dozens of papers, international evaluations, comparison exercises, reference materials and best-practice guides.

By contrast, most of the latest air-pollution sensors are developed by small- and medium-sized enterprises, backed by venture capital and crowdsourced funding. Many devices adapt off-the-shelf technologies. Peer review and academic evaluation may be bypassed. The public are the early adopters; research chemists and physicists are largely on the sidelines. Academics' funding is threatened by this commercial acceleration, because these devices mean that incremental research developments — such as the miniaturization of high-quality detectors, often based on optical absorption, particle counting or mass spectrometry — are less attractive to granting agencies. Many of the processes used for cheap sensors, such as chemical interactions between gases and surfaces, are less well understood.



A sensor used to measure air quality in Kenya.

ALEXANDER IKAWAH

The range of devices is wide. The cheapest, costing a few dollars each, use technologies that have been repurposed from hazard detectors, such as metal-oxide sensors that measure oxidizable gases. For tens to hundreds of dollars, electrochemical or photoionization detection can notionally observe particular compounds or classes. In the $150–1,500 band come miniaturized instruments, such as optical particle counters that can fit in your palm. In general, reducing cost inevitably reduces specificity or sensitivity, or both.

KEEP TESTING

Most commercial sensors target parameters that governments need to track, such as levels of particulate matter (PM) and NO_2. To do a thorough job requires calibration of the target compound and all other possible interferences that might be present. City authorities and the public lack the technical means of checking these themselves, so must take the quality of the measurements on trust from the supplier. The US Environmental Protection Agency has created a technical framework for testing sensors in public use, benchmarking them against the most accurate monitors. But manufacturers might not engage with this process unless they are required to.

The literature on real-world sensor performance is thin. Anecdotally, we have heard that leading research labs have tested commercial sensors and found them wanting. But because papers reporting negative results have low priority, only a few studies have been published (see, for example, refs 5 and 6). These reveal stability and sensitivity issues, and show that the sensors react to other air pollutants and longer-lived gases such as CO_2 and hydrogen. They are also influenced by meteorological conditions such as humidity, temperature and wind speed.

Simple sensors perform best when pollution levels are high and when the compound of interest swamps others — for example, sensors for nitric oxide (NO) and NO_2 seem to work well in locations that have heavy traffic and high pollution levels, where concentrations of these gases approach the parts-per-million level. In more typical conditions, sensors might respond to other atmospheric species as well. Calibrations of cheap sensors performed in the lab and in the field can differ markedly[3], and most relationships observed in the field only apply to that location and for a limited time.

Our research shows that the biggest headaches are caused by interfering chemicals, such as CO_2 and H_2, and by the irreproducibility of measurements. Our real-world test of 20 identical ozone sensors on a roof found a difference of a factor of 6 between the highest and lowest measurements[7]. In other words, the variability of the responses was greater than that of the actual atmosphere. We tested a mid-priced electrochemical sensor for NO_2 in real conditions for an atmospheric concentration of 40 micrograms per cubic metre (the European air-quality limit value). We found that roughly half of the signal from the sensor was from NO_2, and that the rest came from the sensor's response to ambient CO_2. The device was detecting changes in air pollution minute by minute, but not only changes in NO_2.

FAIR USE

Does it matter that a sensor reports an indicative value or trend? It depends on how they are sold and used. Some cheap devices are advertised as being simply for raising awareness of pollution, and it might be expecting too much for them to report accurate values. Others claim to give pollutant measures that can be compared against conventional monitors or official model forecasts.

Until there is agreement on what degree of sensor accuracy is acceptable, we urge caution. Their fitness for purpose should be demonstrated, particularly where they will have a role in decision-making — whether it is at a city, community or personal level. Although we do not wish to stifle innovation, sensors that claim to be able to measure ambient pollution levels could be required to undergo an independent testing regime, as is the case for instruments that are used in regulatory measurements. Some definition of measurement

© 2016 Macmillan Publishers Limited. All rights reserved.

uncertainty is needed, as is standard practice in other fields — even bathroom scales come with uncertainties printed on them. A mark should signify that the sensor meets a minimum quality standard

If such a stamp of approval sounds bureaucratic, think of how the data might be used. People with asthma might use their local sensor data to make personal decisions on medication; an air-pollution sensor is not meant as a medical device, but its real-world application could make it function like one. Privately owned sensor data could trigger legal actions in areas that apparently exceed local air-quality standards. The economic and socially disruptive costs of closing roads or banning cars based on live sensor data would be huge.

NEXT STEPS

The academic air-pollution community must do the hard yards in the lab and field on calibration and testing. It must also find ways to overcome some measurement challenges. Researchers should take the lead on evaluating sensor performance, creating better devices and designing research applications that are suited to the quantified capabilities of sensors.

More creativity is needed in experimental design. If the long-term performance of sensors is a problem, as is likely, then we need to design shorter-term experiments that can be performed reliably. For example, a fine-scale but qualitative measure of pollution might help to simulate the turbulent flows of pollution in street canyons or tree canopies over a few days. There might be experiments in which a fast-responding bulk sensor — one that measures the sum of many organic compounds, for example — might be able to track rapid temporal changes that add context to a slower but more quantitative instrument, such as a gas chromatograph or diffusion tube. Statistical and machine-learning methods might be developed to enable better extraction of signals from a mix of pollutants[8].

"Manufacturers and regulators need to define how and where sensors can be used."

However, academics should not become gatekeepers or validation bodies. This is a job for manufacturers and regulators, who need to define how and where sensors can and cannot be used effectively.

Governments must provide advice now to potential 'professional users', such as in cities and regional environmental agencies. For sensors that might be used for public policy, health studies or any type of infrastructure control, independent testing and verification is essential, as is already being done through long-standing environment-agency committees and national air-pollution schemes. Even sensors that are designed for entertainment or awareness-raising need appropriate labelling to define their capabilities.

Well designed sensor experiments, that acknowledge the limitations of the technologies as well as the strengths, have the potential to simultaneously advance basic science, monitor air pollution — and bring the public along. ■

Alastair Lewis *is a science director at the National Centre for Atmospheric Science in Leeds, UK, and professor of atmospheric chemistry at the University of York, UK.*
Peter Edwards *is a research fellow in the Wolfson Atmospheric Chemistry Laboratories at the University of York, UK.*
e-mails: ally.lewis@ncas.ac.uk; pete.edwards@york.ac.uk

1. Lim, S. S. *et al. Lancet* **380,** 2224–2260 (2012).
2. Kumar, P. *et al. Environ. Int.* **75,** 199–205 (2015).
3. Piedrahita, R. *et al. Atmos. Meas. Tech.* **7,** 3325–3336 (2014).
4. Nieuwenhuijsen, M. J. *et al. Environ. Sci. Technol.* **49,** 2977–2982 (2015).
5. Mead, M. I. *et al. Atmos. Environ.* **70,** 186–203 (2013).
6. Kamionka, M., Breuil, P. & Pijolat, C. *Sens. Actuators B Chem.* **118,** 323–327 (2006).
7. Lewis, A. C. *et al. Faraday Discuss.* http://dx.doi.org/10.1039/C5FD00201J (2015).
8. De Vito, S., Piga, M., Martinotto, L. & Di Francia, G. *Sens. Actuators B Chem.* **143,** 182–191 (2009).

Make peer review scientific

Thirty years on from the first congress on peer review, **Drummond Rennie** reflects on the improvements brought about by research into the process — and calls for more.

Peer review is touted as a demonstration of the self-critical nature of science. But it is a human system. Everybody involved brings prejudices, misunderstandings and gaps in knowledge, so no one should be surprised that peer review is often biased and inefficient. It is occasionally corrupt, sometimes a charade, an open temptation to plagiarists. Even with the best of intentions, how and whether peer review identifies high-quality science is unknown. It is, in short, unscientific.

A long time ago, scientists moved from alchemy to chemistry, from astrology to astronomy. But our reverence for peer review still often borders on mysticism. For the past three decades, I have advocated for research to improve peer review and thus the quality of the scientific literature. Here are some reflections on that winding, rocky path, and some thoughts about the road ahead.

I trained as a physician, studying the pathophysiology of exposure to high altitudes. In 1977, I became deputy editor of *The New England Journal of Medicine* (*NEJM*), working with what I assumed was a smoothly oiled peer-review system. I found myself driving an enormous machine whose operation was sometimes interrupted by startling hiccups. The first big one occurred a year after I arrived. An author who had submitted a paper to our journal accused one of our reviewers, who worked at a competing lab, of plagiarizing parts of her paper. She sent us a manuscript that her lab chief had been sent to assess for another journal, one that I could see had been typed on the same typewriter that the reviewer had used to write his review. I was told to sort it out.

This was more than a decade before a formal definition of research misconduct and systems for its investigation were established. Several careers fell apart. That of the actual plagiarist, and also that of his chief, our reviewer, who was the senior co-author of the manuscript that contained the plagiarism. Tragically, our innocent submitting author also gave up research when her accusations were rebuffed, and she was bullied and demeaned for her persistence and integrity.

This slow-motion catastrophe angered me. How common was such incompetence, confusion and corruption? Did peer review root it out — or just lob it down the road? A few years later, revelations of fabricated data in scores of papers by US cardiologist John Darsee, in *NEJM* and other journals, showed that peer review was usually helpless in detecting gross fraud. More recently, the cases of Dutch psychologist Diederik Stapel and US-based cancer researcher Anil Potti underline how easily false data continue to get through the system. Even if peer review could not detect outright

© 2016 Macmillan Publishers Limited. All rights reserved.

SELECTING GOOD SCIENCE

Milestones in modern peer review and reporting.

1978–79 Revelations of scientific fraud at Yale and Harvard universities publicizes the issue.

1978–92 The Oxford Database of Perinatal Trials is set up by Iain Chalmers. He later establishes the Cochrane Collaboration and its systematic analyses.

1986 Studies demonstrate publication bias in clinical trials; it is caused by the failure of trial authors to submit results for publication.

1989 Regulations defining scientific misconduct and a procedure to address allegations are codified into US law. Peer review is revealed to be ineffective against misconduct.

1989 The first Peer Review Congress held in Chicago, Illinois. It includes a trial of blinding reviewers to authors' identities.

1993 The Cochrane Collaboration, founded to review published reports relevant to health, reveals inherent biases.

1996 The CONSORT statement on reporting clinical trials is released, with a checklist to assist authors and reviewers.

1999 The *British Medical Journal* adopts open peer review on the basis of evidence from randomized trials of the practice.

2000–PRESENT Online-only journals rise in prominence along with new models of peer review.

2004 Clinical-trial pre-registration is made a condition of publication.

2006 The EQUATOR Network is founded to assemble reporting guidelines.

2010 'Beall's list' warns against 'predatory' journals with questionable peer review.

2014–PRESENT Groups (including ORCID, CASRAI, F1000 working group) are founded to support and credit reviewers.

2017 Eighth Peer Review Congress to be held in Chicago.

fabrications, could it sniff out error in honest scientific work, I wondered? There had to be a way to find out.

QUESTIONS ASKED

In 1985, an influential commentary[1] asserted that "the arbiters of rigor, quality, and innovation in scientific reports" did not "apply to their own work the standards they use in judging the work of others". Ouch! Peer review had to be studied, it said, and the most urgent need was leadership within the scientific community.

I had been working at *The Journal of the American Medical Association* (*JAMA*) since 1983. The chief editor was interested in holding a conference on peer review; I jumped at the chance. I insisted that all presentations describe research — and then worried whether we would get a single abstract.

The inaugural Peer Review Congress was held in a distinctly shabby hotel in Chicago, Illinois, in 1989. It was engaging and contentious: presenters studied the demography of reviewers at various journals, how often individuals conducted reviews, blinding, statistical reporting and much more. I was thrilled to see actual data.

A distinguished editor in the audience took another view, excoriating presentation after presentation. Finally, Iain Chalmers (who later co-founded the Cochrane Collaboration) stood and addressed him: "We have listened to your incessant criticisms of everyone who has gone to the trouble of obtaining data. What we have not heard from you is one single piece of evidence for your opinions." There was loud applause, and the future of these congresses was assured. They have taken place every four years since — in much better hotels.

Thanks to such research, we now know a great deal about the mechanics of peer review — the time taken to appraise papers, rates of disagreement between reviewers, the cost at certain journals, even the occurrence of misconduct during review.

Research has brought clear improvement to the biased reporting of clinical trials. Randomized clinical trials cost millions of dollars, are rarely repeated, and greatly influence what treatments patients receive. My colleagues and I showed that most trial results in submitted manuscripts favoured the treatment tested, and this was reflected in the results that were published[2]. Other work revealed that more than 90% of the bias was due to authors failing to submit manuscripts that are unfavourable to the treatment, and that commercial sponsorship drove decisions not to submit[3]. Although any single trial might have been conducted well, the system was skewed. Publication bias made drugs look better than they were.

This line of investigation provided evidence that convinced journals to require that clinical trials be 'pre-registered' at inception. Compliance is still patchy, but journal editors now routinely check that trials were announced publicly (typically at ClincialTrials.gov) before results were collected. We can now expect that when drugs are found to cause serious harm during the trials, the existence of those trials will no longer be hidden from the world.

Meta-research has revealed other sources of distortion. For instance, when trial reports fail to account for control patients or do not fully describe methods for randomization and blinding, they are also more likely to report exaggerated effects.

Such observations led to new standards for reporting clinical trials. An early version of the guidelines was tested in *JAMA* and produced a report that our readers found unreadable[4]. The next version of the guidelines, published in 1996 and called CONSORT (Consolidated Standards of Reporting Trials, of which I am a co-organizer), was much better accepted. These proved a highly successful model for reporting, say, epidemiologic studies, or reports of assessing clinical tests[5]. A collection of more than 300 reporting guidelines have been gathered into the EQUATOR Network (www.equator-network.org), and their use is spreading widely among biomedical researchers, journals and reviewers.

Meta-research on clinical trials has been further advanced by the Cochrane Collaboration, which systematically collects studies across disease types to weigh up the evidence. Cochrane has developed 'risk of bias' assessments to help its reviewers to evaluate possible weaknesses in trial reports.

OPEN REVIEW

Blinding of reviews is another fertile area of study. In 1998, my colleagues and I conducted a five-journal trial[6] of double-blind peer review (neither author nor reviewer knows the identity of the other). We found no difference in the quality of reviews. What's more, attempts to mask authors' identities were often ineffective and imposed a considerable bureaucratic burden. We concluded that the only potential benefit to a (largely unsuccessful) policy of masking is the appearance, not the reality, of fairness. Since then, online technologies for blinding have increased, as have numbers of scientists (and thus the difficulty of guessing who authors may be). It will be interesting to see how similar studies work out now, and whether double-blind reviewing affects acceptance rates for women and under-represented minorities.

More than a decade ago, the *British Medical Journal* (*BMJ*) ran trials in which the identities of both author and reviewer were disclosed to each other during review, and, if the paper was published, the reviewers' names were made public. The *BMJ* did not suffer a loss of manuscripts or reviewers, and

© 2016 Macmillan Publishers Limited. All rights reserved.



ILLUSTRATION BY DAVID PARKINS

now makes such disclosures compulsory. Its experience suggests that how questions are posed is crucial. If a survey asks: "Would you like to sign your review?", most will decline. But if an editor says: "Our journal requires signed reviews. Will you review?", the *BMJ*'s experience is that very few will refuse[7]. I believe that this brand of open review is the most ethical variety, and its practicability is established. In the present system, authors frequently misidentify reviewers with complete confidence, so blame falls on innocent bystanders.

THE FUTURE

The past 15 years have seen an exciting surge of experimentation with new models of peer review — open, blinded, pre- and post-publication, portable and so on[8]. Some of these systems were tried and abandoned decades ago, before the Internet eased testing and logistics.

We need rigorous studies to tell us the pros and cons of these approaches today. Until then any advertised advantages of new arrangements are unsupported assertions. A 2015 survey[9] of more than 1,000 manuscripts was encouraging about the ability of review to identify important papers, but still found lapses.

After all, online technologies don't give reviewers more time or stamina. A common claim of new journals, whether legitimate or 'predatory' (those that charge fees to publish, but that do not offer standard publishing services), is rapid review and publication. This is a powerful pull for authors, but the detailed attention and mature reflection required for a constructive review takes time.

So what now? In my field, and perhaps in many others: follow the triallists. First, develop evidence-based lists of items to be included in reporting (mission-sort-of-accomplished for many clinical journals). Journals must accept and promote these guidelines and ensure that reviewers hold authors to them; perhaps they should facilitate training in peer review, which has been shown to improve performance. Finally, manuscript editors and copy editors must uphold the standards. For example, we now routinely reject trial reports that cannot prove registration before inception. This change is large for all involved — authors, reviewers and journal staff — and it is taking years.

And we must continue to study what we have done. Assessment of review is more likely now than ever before. The two-year-old Meta-Research Innovation Center (METRICS) Institute at Stanford University in California, which is devoted to researching and improving the process of science, shows that the field is maturing and gaining respect. So does last year's launch of the journal *Research Integrity and Peer Review*, a home for research on the topic.

> ***"We need rigorous studies to tell us the pros and cons of these approaches."***

In 1986, we were lucky with our timing. The peer-review congresses came just as others were trying to see what could be learned from the literature to arrive at the best treatments for patients, developing methods for systematic review, and nailing down the biases that pervade clinical research (see 'Selecting good science'). These people did the work.

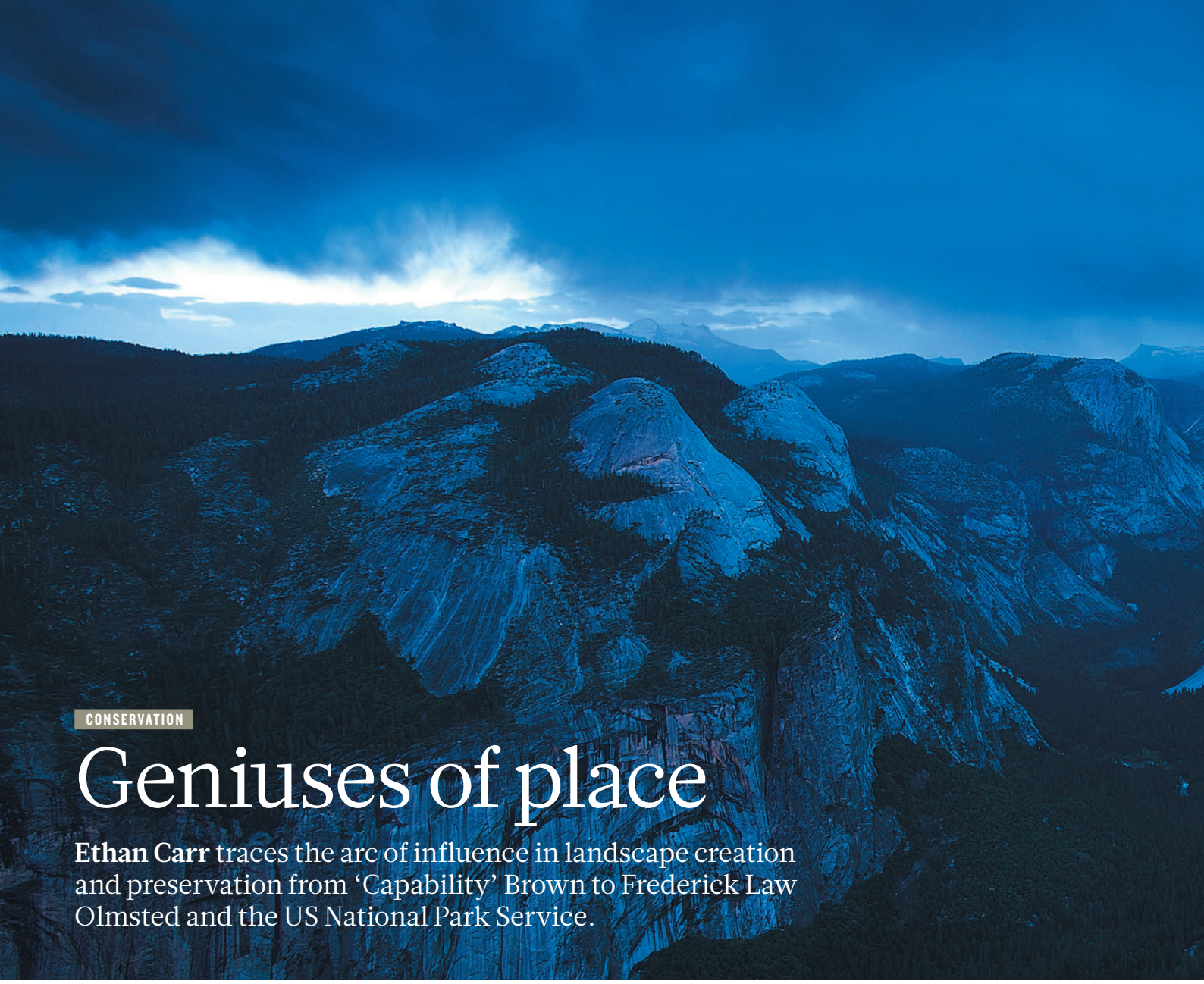
To announce that first Peer Review Congress, I wrote: "There are scarcely any bars to eventual publication. There seems to be no study too fragmented, no hypothesis too trivial, no literature citation too biased or too egotistical, no design too warped, no methodology too bungled, no presentation of results too inaccurate, too obscure, and too contradictory, no analysis too self-serving, no argument too circular, no conclusions too trifling or too unjustified, and no grammar and syntax too offensive for a paper to end up in print"[10].

Unfortunately, that statement is still true today, and I'm not just talking about predatory journals. That said, I am confident that the Peer Review Congress scheduled for 2017 will be asking more incisive, actionable questions than ever before. ■

Drummond Rennie *is a co-organizer of CONSORT, a former member of the Commission on Research Integrity for the US Public Health Service, and former president of the World Association of Medical Editors.*
e-mail: drummond.rennie@ucsf.edu

1. Bailar, J. C. & Patterson, K. *N. Engl. J. Med.* **312,** 654–657 (1985).
2. Olson, C. M. *et al. J. Am. Med. Assoc.* **287,** 2825–2828 (2002).
3. Dickersin, K. & Rennie, D. *J. Am. Med. Assoc.* **290,** 516–523 (2003).
4. Rennie, D. *J. Am. Med. Assoc.* **273,** 1054–1055 (1995).
5. Begg, C. *et al. J. Am. Med. Assoc.* **276,** 637–639 (1996).
6. Justice, A. C. *et al. J. Am. Med. Assoc.* **280,** 240–242 (1998).
7. Groves, T. *Br. Med. J.* **341,** c6424 (2010).
8. Paglione, L. D. & Lawrence, R. N. *Learn. Publ.* **28,** 309–316 (2015).
9. Siler, K., Lee, K. & Bero, L. *Proc. Natl Acad. Sci. USA* **112,** 360–365 (2015).
10. Rennie, D. *J. Am. Med. Assoc.* **256,** 2391–2392 (1986).

© 2016 Macmillan Publishers Limited. All rights reserved.



CONSERVATION

Geniuses of place

Ethan Carr traces the arc of influence in landscape creation and preservation from 'Capability' Brown to Frederick Law Olmsted and the US National Park Service.

JEFF MURRAY/AURORA

A coincidence of commemorative dates makes this year an important one in the history of landscape design and scenic preservation. As the 300th anniversary of the birth of the landscape gardener Lancelot 'Capability' Brown is celebrated on one side of the Atlantic, the United States is marking the centenary of the National Park Service, the federal agency that acts as the steward of the nation's most iconic natural areas and historic shrines. The two are connected by the complex and evolving cultural construction of 'nature', its representations, its manifestations and its benefits.

Brown's landscape parks expressed the eighteenth-century's fascination with nature itself, which was increasingly the subject of scientific inquiry and a plethora of botanical and zoological discoveries. Nature offered templates for ordering society, too. When the poet Alexander Pope exhorted, "In all, let Nature never be forgot," he was describing more than the new style of landscape gardening. Brown's composed scenes of pastoral greenswards and planted woodlands expressed picturesque aesthetic theory; they also imposed a more scientific and modern order on the land.

In the United States, landscape architect Frederick Law Olmsted developed his own 'natural style' in the nineteenth century. Olmsted was deeply influenced by his experiences in Britain, which he described in his first book, *Walks and Talks of an American Farmer in England* (1852). In the spring of 1850, he visited Birkenhead Park in northwest England, noting that in "democratic America there was nothing to be thought of as comparable to this People's Garden". Olmsted also responded to the countryside itself, and, above all, to the landscape parks he visited. About the designer of the grounds at Eaton Hall in Cheshire, he wrote: "What artist, so noble … as he who, with far-reaching conception of beauty and designing power, sketches the outline, writes the colours, and directs the shadows of a picture so great that Nature shall be employed upon it for generations, before the work he has arranged for her shall realize his intentions."

That artist was Brown, who had died in 1783. His park landscapes, now mature, thoroughly impressed the young "American farmer". Sweeping meadows, clumps and belts of native (and North American) trees, sheets of impounded water and winding drives were the elements that shaped the aesthetics and image of the "natural" in an urbanizing and industrializing world.

Olmsted soon assumed the mantle of artist himself. At first, he worked with a partner: the English architect Calvert Vaux, with whom he won the competition for the design of New York City's Central Park in 1858. Olmsted returned to England several times. He expressed ambivalence towards Victorian design trends, which emphasized 'gardenesque' displays of floriculture and frankly artificial arrangements of

NATURE.COM
For more on science in culture see:
nature.com/booksandarts

© 2016 Macmillan Publishers Limited. All rights reserved.



California's Yosemite Valley was one of the first US national parks.

specimen plants. Olmsted's taste could be considered atavistic. Like Brown, he sought to create compositions of larger 'landscape effects', devoid of elaborate flower gardens or other distractions from the fundamental experience of scenery. In practice, Olmsted created dramatic sequences of landscapes — expansive greenswards, serpentine lakes, picturesque rambles — and eschewed buildings, geometric layouts and flower beds. He would not have another kindred spirit in British landscape gardening until 1870, when William Robinson — who later designed the grounds of Gravetye Manor in Sussex — published the book *Wild Garden*. Robinson visited Olmsted in New York that year, and the two maintained a correspondence and a mutual admiration.

THE LIE OF THE LAND

Olmsted was also influenced by continued progress in contemporary natural sciences, especially geology, which he knew mostly through the work of the researchers Louis Agassiz and Nathaniel Shaler at Harvard University in Cambridge, Massachusetts. With Vaux and on his own, Olmsted exploited existing geological formations in his large municipal-park designs to create specific effects and to structure the overall landscape composition.

> ***"Nature offered templates for ordering society."***

The schist bedrock outcrops of Manhattan and the puddingstone conglomerate of Boston, Massachusetts, are design features (and construction materials) in Central Park and Franklin Park, respectively. In Brooklyn, New York, the terminal moraine glacial morphology of Long Island became the framework for the entire conception of Prospect Park as a sequence of landscape experiences, from the high ground of the main entrance down to the glacial outwash plain, in which a large lake was excavated. What Pope described as the "genius of the place", for Olmsted, resided in the landscape's skeleton — its geological foundations — which he often exposed and highlighted, and around which landscape effects and overall patterns of how people might use the park could be structured.

As a public intellectual, Olmsted also developed the political rhetoric and economic justifications for larger regional and national scenic reservations. In 1865, the governor of California asked him to prepare a report to guide the management of Yosemite Valley. This granite gorge hidden in the Sierra Nevada mountains is one of the great geological landscapes of the continent. It became the site, more than any other, where the idea of the national park took shape. It was for Yosemite that Olmsted provided the philosophical framework for state and national park-making in the United States. He noted that it was "the main duty of government" to protect and provide the means for the "pursuit of happiness". That pursuit, for Olmsted, ▶

© 2016 Macmillan Publishers Limited. All rights reserved.

depended on preserving such places and creating public access to them.

"It is a scientific fact," he asserted, "that the occasional contemplation of natural scenes of an impressive character … is favorable to the health and vigor of men." The government had a duty to assure that "enjoyment of the choicest natural scenes in the country and the means of recreation connected with them" be "laid open to the use of the body of the people". If the government did not act, those places would be monopolized by the few and their benefits experienced only by an elite. The establishment of "great public grounds" was therefore required of a republic that derived its authority from its people.

There was a continuity and consistency in the overall purposes that Olmsted described for public parks and scenic reservations, as well as in his design recommendations for both. At Yosemite Valley and New York's Niagara Falls (for which he and Vaux prepared the state-park plan in 1887), for example, the challenge was to protect the awesome existing features from damage by visitors, and to choreograph the sequence and pace of their visits in the design of roads, paths and other facilities, without marring the scenery with buildings.

SHAPING DEMOCRACY

Brown is supposed to have said, "One does not go up and down steps in nature", referencing his preference for smoothly graded contours over retaining walls or terraces. In their Central Park competition entry, Olmsted and Vaux similarly insisted that: "the interest of the visitor … should concentrate on features of natural, in preference to artificial, beauty … Architectural structures should be confessedly subservient to the main idea." In the changed context of nineteenth-century, urbanizing US society, the main purpose of the large, public park (whether municipal, state or national) remained constant: to provide a dramatic sequence of affecting landscape experiences and effects, unencumbered by encroachments, and now made available to "the body of the people".

This rhetoric of public park-making is particularly important while the centenary of the National Park Service is being celebrated. Congress created the agency in 1916, giving it a famous mandate "to conserve the scenery and the natural and historic objects and the wild life" of the national parks, and "to provide for the enjoyment of the same in such manner and by such means as will leave them unimpaired for the enjoyment of future generations". This key portion of the legislation was written by Frederick Law Olmsted Jr, who continued his father's professional practice in the twentieth century, and who was directly inspired by his father's Yosemite report in drafting the park-service bill.





TOP: OLAF PROTZE/LIGHTROCKET/GETTY; BOTTOM: MICHAEL YAMASHITA/NGC

Top: Capability Brown's garden at Blenheim Palace, UK; bottom: Central Park, New York.

Congress had created national parks in the mid-nineteenth century — notably Yosemite in 1864 and Yellowstone, Wyoming, in 1872. But the far-flung group of about 35 reservations had remained relatively inaccessible to most people. That changed with the advent of affordable and reliable automobiles. The park service was created to better manage both the great potential for public enjoyment and the great peril to the parks presented by vastly increased numbers of tourists in cars.

Today, there are more than 400 'units' in the national-park system, including scores of historic sites, memorial landscapes and archaeological sites, in addition to the better-known large-scale wilderness reservations. The national parks are often characterized as 'America's best idea', a bromide that obscures as much as acknowledges their significance and origins. The idea was rooted in the nineteenth-century park movement, and therefore in the thought reflected in the elder Olmsted's writings, and embodied in his designs. These in turn have unambiguous links to that "artist so noble", born 300 years ago, Capability Brown. ■

Ethan Carr *is a landscape historian and preservationist specializing in public landscapes at the University of Massachusetts Amherst. He is the author of* Wilderness by Design *and* Mission 66, *and the volume editor of* The Papers of Frederick Law Olmsted, Volume VIII: The Early Boston Years, 1882–1890.
e-mail: ecarr@umass.edu

© 2016 Macmillan Publishers Limited. All rights reserved.

Correspondence

Stop slaughter of migrating songbirds

A new strategy is needed to stop the illegal trapping and killing of millions of songbirds every year in the Mediterranean region, where gigantic vertical nets intercept major migration flyways (see also *Nature* **529,** 452–455; 2016). In the western Mahgreb in North Africa, this carnage is collateral damage to the area's cultural fancy for pet goldfinches (*Carduelis carduelis*), which dates back to around 700 and the Umayyad dynasty.

The goldfinch has only recently been officially protected in Algeria, Tunisia and Morocco, where its populations have been declining rapidly over the past two decades. The price of a single live bird (pictured) is now US$50–500, equivalent to 25–250% of the typical local monthly salary. This has caused trapping and by-catch to escalate. Many of the captured goldfinches perish under poor transport conditions.

I suggest that local people should be taught to divert their admiration for the goldfinch's charms into ensuring its protection. Netting would stop if instead the goldfinch became an emblematic conservation symbol of the region, and a 'safety umbrella' for other migrating Palaearctic songbirds (see J.-M. Roberge and P. Angelstam *Conserv. Biol.* **18,** 76–85; 2004).

Rassim Khelifa *Université Mouloud Mammeri de Tizi Ouzou, Algeria.*
rassimkhelifa@gmail.com



Goldfinches for sale in an Algerian market, despite being officially protected.

Don't undervalue the social sciences

Too many physicists, chemists and biologists perceive the social sciences and humanities as less rigorous and less intellectually demanding domains than their own. Research into expert performance calls these attitudes into question.

Thousands of hours of deliberate practice are needed to become highly competent in any endeavour that requires skill (see *The Cambridge Handbook of Expertise and Expert Performance* Cambridge Univ. Press; 2006). Moreover, the time invested before making a world-class contribution to any major field is similar, be it in chess, music, basketball, history or flying a plane (A. Ericsson and R. Pool *Peak* Bodley Head; 2016).

So, distinguished scholars from different fields are likely to be comparably proficient in the skills relevant to their work. Assuming that top researchers have devoted roughly the same amount of effort to developing their domain-specific skills, the wider implication is that different fields are roughly equally advanced in terms of dealing with the challenges they face.

If physics, say, seems more developed than social science, then this may be because the field has been established for longer or that the challenges are easier to overcome.

Brian Martin *University of Wollongong, Australia.*
bmartin@uow.edu.au

US panel risks infant and researcher lives

As the chief executives of the biotech companies Ganogen and StemExpress, we are among a broad sweep subpoenaed — along with scientists, graduate students and physicians also engaged in research involving fetal tissue — by the US House Select Investigative Panel on Infant Lives. In our view, this witch-hunt endangers infants and researchers and must end.

The panel's stated aim is to "get the facts about medical practices of abortion service providers and the business practices of the procurement organizations who sell baby body parts". On 1 June, it released the names, addresses, e-mail contacts and telephone numbers of many of us in an open letter to the US Department of Health and Human Services. We consider this to be a callous disregard of the threat posed by activists to medical researchers who are in fact engaged in saving young lives (see *Nature Biotechnol.* **34,** 445; 2016).

Research involving fetal tissue led to vaccines against polio, rubella and chickenpox. It was central to proving the link between Zika virus and infant microcephaly (H. Tang *et al. Cell Stem Cell* **16,** 587–590; 2016), and is essential for developing a vaccine against the virus (*Nature* **532,** 16; 2016). The chair of the panel, Representative Marsha Blackburn of Tennessee, should note that her constituents, and those of committee members Diane Black (Tennessee) and Vicki Hartzler (Missouri), are especially vulnerable to Zika because the mosquito vector, *Aedes aegypti*, is more prevalent in the southern states.

Eugene Gu *Ganogen Research Institute, Redwood City, USA.*
Cate Dyer *StemExpress, Placerville, USA.*
eugenegu@ganogen.org

Food security needs social-science input

As members of the Climate-Resilient Open Partnership for Food Security project supported by the World Wide University Network (see go.nature.com/28ygwtc), we contend that basic social-science theory and methods could transform interventions aimed at improving food production.

Food security calls for agricultural advances, adaptation to climate change and more efficient use of natural resources. Just as important are the social and political considerations of reforming food production and distribution systems.

All too often, poor communication between the scientific community and the public, including potential users, impedes utilization of new technologies. Social networks, power inequalities and institutional resistance to change must all be taken into account if the system is to be reformed (see W. W. Powell *et al.* in *The Science of Science Policy* 31–55, Stanford Univ. Press; 2011).

We therefore suggest that research consortia in food security and their funding agencies should include social scientists from the outset (see A. Viseu *Nature* **525,** 291; 2015). This would dramatically enhance project management and conceptual development by dealing with the complex interactions between natural and social factors.

Klaus Nüsslein*, Om Parkash Dhankher* *University of Massachusetts Amherst, USA.*
nusslein@microbio.umass.edu
**Supported by 13 signatories (see go.nature.com/299szyy).*

© 2016 Macmillan Publishers Limited. All rights reserved.

ASTROPHYSICS

Rare data from a lost satellite

The Hitomi astronomical satellite observed gas motions in the Perseus galaxy cluster shortly before losing contact with Earth. Its findings are invaluable to studies of cluster physics and cosmology. SEE LETTER P.117

ELIZABETH BLANTON

Measurements of motions in the hot gas that lurks in clusters of galaxies provide insight into the level of turbulence in these objects, and into larger-scale flows that are related to a cluster's merger history and to outflows from its central supermassive black hole. The amount of turbulence affects measurements of cluster mass, which are used to constrain values of the cosmological parameters that govern our Universe. Indirect methods have been used previously to infer these gas motions, but on page 117, the Hitomi collaboration (Aharonian *et al.*[1]) reports direct measurements of gas motions in the Perseus cluster using high-resolution spectra acquired by a type of detector that was unique to the Japanese Hitomi X-ray astronomical satellite; similar detectors are not available on any other active X-ray satellite. The Perseus cluster was the only cluster to be observed by Hitomi before the satellite's premature demise[2], which means that such observations will be rare for the foreseeable future.

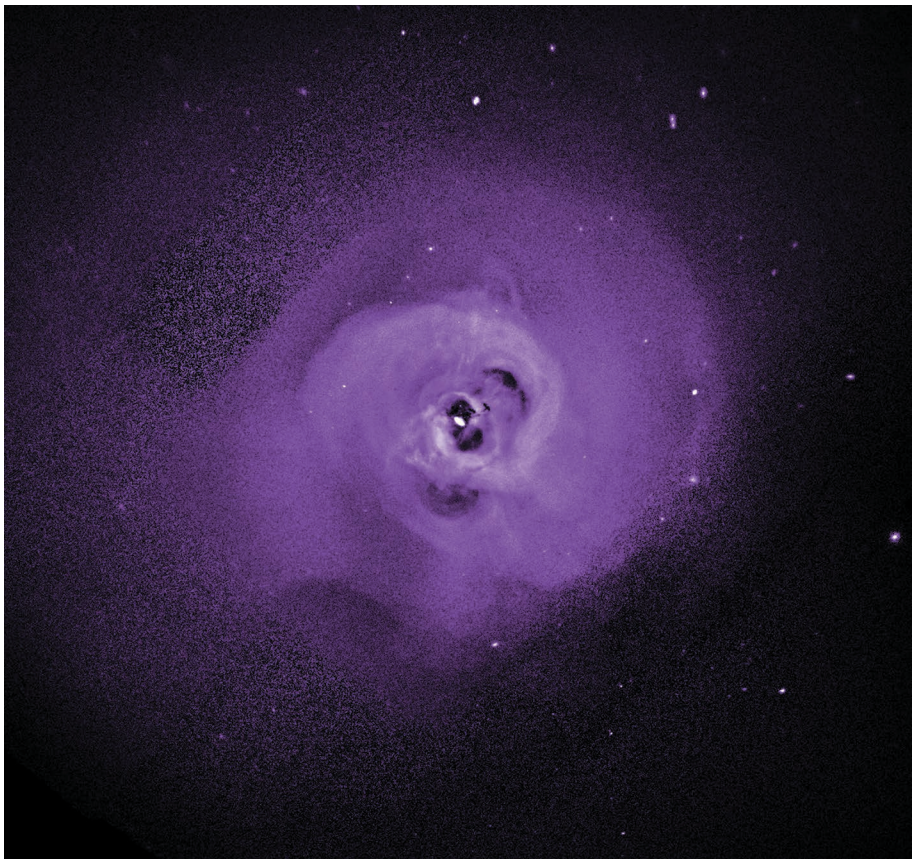
Clusters of galaxies are the largest gravitationally bound objects in the Universe, and are crucial probes of galaxy evolution and of cosmology. In the currently favoured lambda cold dark matter model of the Universe, structure forms in a 'bottom-up' fashion — smaller structures form first and then merge to generate larger ones. Individual galaxies and small groups of galaxies therefore form before merging into more-massive clusters.

Cosmological models vary, for example, by including different amounts of dark matter and dark energy in the Universe, and thus predict different cluster-formation histories. The observed mass distribution of clusters as a function of time over the evolution of the Universe can place constraints on these models[3,4]. But the masses of individual clusters must be measured robustly to place these constraints with high accuracy.

Clusters of galaxies typically contain from about fifty to thousands of galaxies, along with dark matter and diffuse, hot (about 10^7–10^8 kelvin) gas that emits X-ray radiation. One way of measuring a cluster's mass is to analyse the emission from this gas, which can be assumed to be in hydrostatic equilibrium — that is, the pressure gradient of the gas in the outward direction is balanced by gravity pulling inward; this gravity depends directly on the total mass of the cluster. The gas pressure can be readily determined from measurements of gas density and temperature, which are acquired using data from existing X-ray space observatories such as the Chandra X-ray Observatory and XMM-Newton.

But turbulent motions in the gas can potentially add to the pressure, and neglecting their contribution can cause errors in mass determinations. Measuring these motions directly requires high-resolution spectroscopy of the diffuse, hot gas, which can extend across millions of light years in space. Indirect constraints on turbulent gas motions have previously been made using other methods[5,6], and upper limits have been set on the basis of spectroscopic measurements taken by XMM-Newton[7]. The Hitomi X-ray observatory was the only active facility to carry a type of instrument called a calorimeter that could precisely measure motions in the gas through changes in frequency (Doppler shifts) and broadening of emission lines in spectral data.

The Hitomi collaboration chose the Perseus



NASA/CXC/STANFORD/I. ZHURAVLEVA *ET AL.*

Figure 1 | X-ray image of the Perseus cluster of galaxies taken by the Chandra observatory. The central bubbles (dark regions) and ripples are associated with outflows from material that surrounds the cluster's central supermassive black hole. On the basis of measurements taken by the now lost Hitomi satellite, Aharonian *et al.*[1] report that turbulence in the central region of the cluster is low, which suggests that errors in measurements of the masses of galaxy clusters using X-ray observations are small. The image is approximately 460,000 parsecs (1.5 million light years) across.

© 2016 Macmillan Publishers Limited. All rights reserved.

cluster of galaxies as an early observation target because it is the brightest X-ray-emitting cluster in the sky and has been studied extensively by other orbiting X-ray observatories, including Chandra, XMM-Newton and Suzaku[8,9]. It therefore provides an excellent test case for measuring motions in its hot gas. It is considered to be 'relaxed', meaning that it has not undergone a large-scale merger with another massive cluster in billions of years. The central galaxy in the cluster hosts a supermassive black hole, and outflows of high-energy particles orbiting the black hole have inflated large bubbles in the cluster's diffuse gas (Fig. 1). The goals of the study were to measure bulk velocities related to these outflows, as well as turbulence.

The main result is that the velocities of the gas are quite low, approximately 150 kilometres per second. A notable implication of this is that the additional contribution to the pressure that is associated with turbulence is constrained to be only a few per cent of the thermal pressure (the main component of the total pressure). This means that measurements of cluster mass based on X-ray observations of hot gas, assuming hydrostatic equilibrium and neglecting turbulent pressure, will have only small associated errors. This is good news for studies that use the masses as the basis for constraining cosmological parameters[4].

However, these measurements were made for only one cluster and only in the cluster's central region, and are therefore not necessarily applicable to clusters in general. In addition, the observations were made early in the mission, before all of the associated calibration procedures were available. The limitations on the available calibrations translate to an increase in the uncertainties in the measurements, particularly in the systematic errors in the line-of-sight velocities. Nevertheless, even with such uncertainties, these cluster gas velocities are the most precise yet measured.

Future missions — including the European Athena X-ray Observatory, scheduled for launch in 2028, and the possible US X-ray Surveyor mission — should allow further insight into the gas motions in clusters of galaxies. It would be useful eventually to measure velocities in a range of cluster environments, such as cluster cores and outer regions, clusters with and without bubbles that result from outflows around supermassive black holes, and clusters in various stages of mergers. The Perseus observations from Hitomi have given us an important first look at the gas motions in a galaxy cluster, but many more exciting environments and details remain to be explored. ■

Elizabeth Blanton *is at the Institute for Astrophysical Research and the Department of Astronomy, Boston University, Boston, Massachusetts 02215, USA.*
e-mail: eblanton@bu.edu

1. The Hitomi collaboration. *Nature* **535,** 117–121 (2016).
2. Witze, A. *Nature* **533,** 18–19 (2016).
3. Allen, S. W., Evrard, A. E. & Mantz, A. B. *Annu. Rev. Astron. Astrophys.* **49,** 409–470 (2011).
4. Vikhlinin, A. *et al. Astrophys. J.* **692,** 1060–1074 (2009).
5. Zhuravleva, I. *et al. Mon. Not. R. Astron. Soc.* **450,** 4184–4197 (2015).
6. Randall, S. W. *et al. Astrophys. J.* **805,** 112 (2015).
7. Sanders, J. S., Fabian, A. C. & Smith, R. K. *Mon. Not. R. Astron. Soc.* **410,** 1797–1812 (2011).
8. Fabian, A. C. *et al. Mon. Not. R. Astron. Soc.* **418,** 2154–2164 (2011).
9. Simionescu, A. *et al. Astrophys. J.* **757,** 182 (2012).

NEUROSCIENCE

In search of the memory molecule

The protein PKM-ζ has been proposed to regulate the maintenance of memory in rodents, but this theory has been questioned. The finding that another isoform of the protein acts as a backup if PKM-ζ is lacking will influence this debate.

PAUL W. FRANKLAND & SHEENA A. JOSSELYN

An understanding of memory has long been a goal of neuroscience. One question that has attracted particular attention is whether there is a specific molecule that maintains memories. After almost two decades of careful work, neuroscientist Todd Sacktor and colleagues thought they had the answer. In 2006, the authors reported[1] that an atypical isoform of the enzyme protein kinase C, called PKM-ζ, was involved in maintaining memories in mice, and that an inhibitor of PKM-ζ could erase memories. The results were subsequently questioned[2,3], and controversy ensued. Writing in *eLife*, the same group that performed the 2006 study opens a new chapter in this debate[4], arguing that PKM-ζ should be restored to its pre-eminent status as the memory molecule.

More than half a century ago, the psychologist Donald Hebb proposed that the synaptic connections between two neurons are strengthened when the neurons fire together[5]. He suggested that this form of synaptic strengthening provided the basis for the formation of long-term memories, enabling many neurons to be linked together in cell assemblies that serve as the physical substrates of memory, called engrams. It was later discovered[6] that high-frequency neural stimulation led to persistent increases in synaptic strength, known as long-term potentiation (LTP). Most neuroscientists embraced the idea that understanding LTP was the key to understanding memory[7]. The race was on to identify the molecular machinery involved in LTP.

One molecule in particular emerged from the fray. Although dozens of molecules were involved in initial synaptic strengthening following high-frequency stimulation, only PKM-ζ seemed to be crucial for maintaining these strengthened connections[8]. In PKM-ζ, therefore, the activity of a single molecule was linked to the persistence of memory. Subsequently, several experiments showed that inhibition of PKM-ζ after memory formation (for example, by using a 13-amino-acid protein fragment called ZIP, which mimics the natural substrate that inactivates PKM-ζ) led to memory erasure[1,9].

However, enthusiasm surrounding PKM-ζ waned dramatically following the discovery that mice in which PKM-ζ had been deleted showed normal LTP and memory[2,3]. More puzzling still, ZIP produced LTP-reversing and memory-erasing effects in mice that lacked PKM-ζ, similar to its effects in normal mice that expressed the enzyme. The amnesiac effects of ZIP, therefore, must be acting through another mechanism.

Do these results indicate that PKM-ζ is not necessary for memory? Much of the initial clamour surrounding the 2013 papers[2,3] focused on this possibility. It seems unlikely, however, because more than one method for inhibiting PKM-ζ erases memories[9]. An alternative possibility is that PKM-ζ has an essential role in LTP maintenance and memory persistence in normal mice, but compensatory processes that are sensitive to ZIP emerge in PKM-ζ-deficient mice.

This brings us to the detective work of the current study. Tsokas *et al.*[4] first confirmed that ZIP reversed LTP in both normal and PKM-ζ-deficient mice, indicating that trivial procedural differences could not resolve the controversy. Next, the authors showed that induction of LTP produced an increase in PKM-ζ in slices taken from the hippocampal region of the brains of normal mice, and that there was a sustained increase in another atypical protein kinase C isoform, PKC-ι/λ, in slices from PKM-ζ-deficient mice. Moreover, injecting either PKC-ι/λ or PKM-ζ directly into hippocampal CA1 pyramidal neurons induced

© 2016 Macmillan Publishers Limited. All rights reserved.

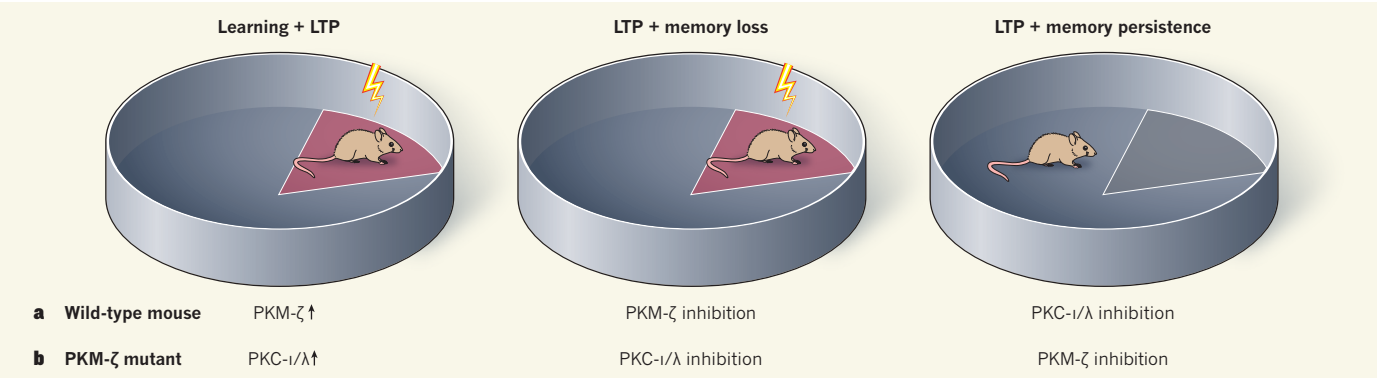


Figure 1 | Memory loss modulated. In place-avoidance tests, mice learn that they will receive a foot shock if they move over a certain part of a rotating test arena. During this learning, the synaptic connections between neurons are strengthened in a process called long-term potentiation (LTP), which is required for memory formation. Tsokas *et al.*[4] investigated how two atypical isoforms of the enzyme protein kinase C — PKM-ζ and PKC-ι/λ — regulate memory maintenance following LTP induction. **a**, In wild-type mice, levels of PKM-ζ rise following learning. Inhibition of PKM-ζ in these mice causes loss of LTP and hence loss of memory, so the mice forget how to avoid a shock. By contrast, inhibition of PKC-ι/λ has no effect on memory of the learned activity. **b**, In mice that lack the gene encoding PKM-ζ, PKC-ι/λ is elevated following LTP induction. Inhibition of PKC-ι/λ causes LTP and memory loss, whereas PKM-ζ inhibition has no effect. Thus PKM-ζ is the main substrate for memory maintenance in normal conditions, but PKC-ι/λ can compensate in its absence.

LTP in slices from normal mice. ZIP treatment reversed the effects of either protein injection, hinting that PKC-ι/λ might be the mystery molecule that compensates for loss of PKM-ζ.

To test this idea directly, Tsokas and colleagues inhibited either PKM-ζ or PKC-ι/λ and examined LTP in hippocampal slices (Fig. 1). In slices from control mice, inhibiting PKM-ζ blocked LTP, but PKC-ι/λ inhibition had no effect. By contrast, in PKM-ζ-deficient mice, inhibiting PKC-ι/λ blocked LTP, but PKM-ζ inhibition was ineffective. The same pattern emerged when the authors examined the effects of PKC-ι/λ and PKM-ζ inhibition on memory in control and PKM-ζ-deficient mice.

Do these latest results restore the position of PKM-ζ as the leading memory molecule? The allure of the PKM-ζ theory is the idea that a single molecule is responsible for maintaining LTP and memories. The current findings are not inconsistent with this view. However, in their experiments, Tsokas *et al.* inhibited PKM-ζ in normal mice before (rather than after) LTP and memory induction. This means that they cannot directly evaluate the enzyme's role in the persistence of LTP and memory.

The PKM-ζ saga serves as a cautionary tale about the specificity of the tools that we use to examine brain function and establish causality. The controversy exposed the bluntness of ZIP as a tool for probing PKM-ζ function because it clearly affects other molecules and may even lead to neuronal silencing[10]. Equally, seemingly more specific interventions, such as genetic deletion of PKM-ζ, produced a cascade of unintended compensatory changes, which clouded interpretations and masked predicted outcomes. This limitation is not restricted to genetic mutations, but extends to any intervention that perturbs brain function (such as optogenetic or chemogenetic strategies in which genetically introduced proteins can be activated and inhibited in response to light or drugs).

As the PKM-ζ debate rumbles on, there is a broader mystery to consider. Molecular neuroscientists such as Tsokas and colleagues present a static view of the engram, in which patterns of synaptic changes that are initiated during memory encoding are maintained over the lifetime of the memory. By contrast, systems neuroscientists present a more dynamic picture, emphasizing memory maintenance in the midst of broad changes in the synapses[11] and even the neurons[12] that correspond to the engram. A full account of memory persistence needs to merge these molecular and systems perspectives, allowing the twain to meet. ■

Paul W. Frankland *and* **Sheena A. Josselyn** *are in the Program in Neurosciences and Mental Health, The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. They are also in the Departments of Physiology and Psychology and at the Institute of Medical Science, University of Toronto, Toronto.*

1. Pastalkova, E. *et al. Science* **313,** 1141–1144 (2006).
2. Lee, A. M. *et al. Nature* **493,** 416–419 (2013).
3. Volk, L. J., Bachman, J. L., Johnson, R., Yu, Y. & Huganir, R. L. *Nature* **493,** 420–423 (2013).
4. Tsokas, P. *et al. eLife* **5,** e14846 (2016).
5. Hebb, D. O. *The Organization of Behavior: A Neuropsychological Theory* (Wiley, 1949).
6 Bliss, T. V. P. & Lømo, T. *J. Physiol. (Lond.)* **232,** 331–356 (1973).
7. Stevens, C. F. *Neuron* **20,** 1–2 (1998).
8. Ling, D. S. F. *et al. Nature Neurosci.* **5,** 295–296 (2002).
9. Shema, R. *et al. Science* **331,** 1207–1210 (2011).
10. LeBlancq, M. J., McKinney, T. L. & Dickson, C. T. *J. Neurosci.* **36,** 6193–6198 (2016).
11. Attardo, A., Fitzgerald, J. E. & Schnitzer, M. J. *Nature* **523,** 592–596 (2015).
12. Rubin, A., Geva, N., Sheintuch, L. & Ziv, Y. *eLife* **4,** e12247 (2015).

This article was published online on 29 June 2016.

CHEMICAL PHYSICS

Quantum control of light-induced reactions

An investigation of how ultracold molecules are broken apart by light reveals surprising, previously unobserved quantum effects. The work opens up avenues of research in quantum optics. SEE LETTER P.122

DAVID W. CHANDLER

The rupture of molecular bonds by the absorption of light drives chemistry in the atmosphere, causes DNA damage and the associated repair response, and provides a superb tool to study how molecules absorb light and then distribute and dispose of its energy. On page 122, McDonald *et al.*[1] report their study of the light-induced break-up (photodissociation) of ultracold strontium molecules, Sr_2. Their work provides insight into how molecules behave in the quantum regime of ultralow-energy dynamics that occurs just above energy thresholds for photodissociation.

Early photodissociation studies focused on

© 2016 Macmillan Publishers Limited. All rights reserved.

the energetics[2] of the products formed from diatomic molecules, and of the products' angular distribution[3] — the distribution of angles at which they recoil relative to the direction of polarization (the polarization axis) of the light that excited them. If the energy of the photon absorbed by the diatomic molecule and the velocities of the resulting atomic fragments were known, then the bond energy of the molecule could be directly determined. The accuracy of these determinations depended on how cold the molecule was initially, and on how accurately one could measure the velocities of the products.

In the early experiments[4], diatomic molecules were irradiated with laser light, and if the fragments were found to fly predominantly parallel to the laser polarization axis, then the transition dipole moment responsible for the light absorption was said to be parallel; similarly, perpendicular transitions were named after the associated perpendicular recoil. The transition dipole moment describes coupling between the two electronic states responsible for light absorption, and this classification was helpful in understanding its nature. For polyatomic molecules, the transition dipole moment does not have to align with a particular molecular axis, and many factors affect the measured angular distribution of the fragments. Measurements of the velocities of fragments provide information about the dynamics of the energy deposited within molecules as it evolves into the kinetic energy of the fragments.

Hundreds of photodissociation studies have been performed because of the fundamental information that can be obtained. With the advent of laser-based imaging techniques[5–8] in the late 1980s, it became possible to measure velocities at high resolution (approximately a few metres per second) for particular electronic states of the products, by projecting the ionized products onto position-sensitive ion detectors. However, these experiments typically used pulsed-dye lasers (which produce light at a low frequency resolution of about 3,000 megahertz) to dissociate molecules and detect the products. This precludes experiments such as those performed by McDonald and colleagues, in which molecules are dissociated by photons that have a much higher, 1 MHz frequency resolution and energies just above the dissociation threshold of the molecule (that is, at light frequencies between 5 and 400 MHz greater than the dissociation-threshold frequency).

Moreover, these experiments typically used supersonic molecular beams as a source of cool molecules. When a high-pressure gas is expanded into a vacuum to form a molecular beam, the flow is directed forward supersonically at the expense of the kinetic energy associated with the other directions of flight and with the gas's internal degrees of freedom (the rotational and vibrational motion of its molecules). This allows molecules to be cooled to temperatures of a few kelvin even though they fly at close to velocities of 1,000 m s^{-1}, with a spread of about 50 m s^{-1}. McDonald and co-workers, however, wanted to study photodissociation fragments moving at only about 1 m s^{-1} (extremely slowly for a molecule, and correlating with a temperature of tens of millikelvin). To see such slow fragments, the authors held their molecules in a stationary laser trap, photodissociated them using a light pulse and then imaged the fragments after they had flown for about a hundred microseconds.

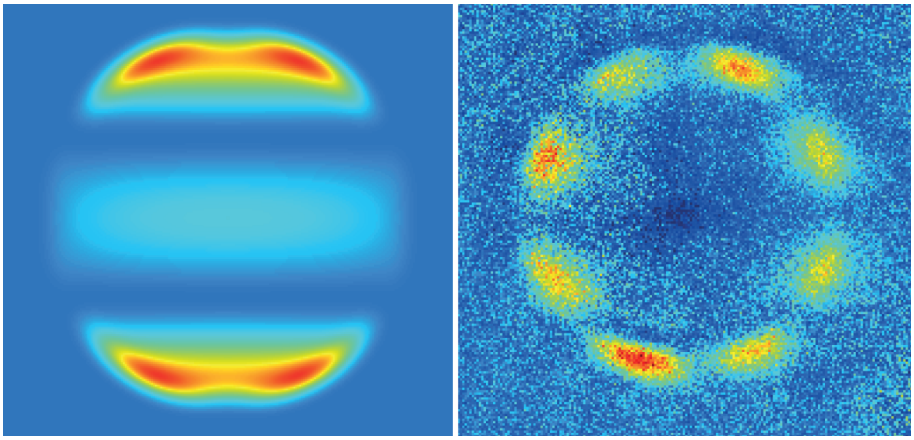


Figure 1 | Quantum effects in photodissociation. McDonald *et al.*[1] studied the light-induced fragmentation (photodissociation) of diatomic strontium molecules, Sr_2, and observed surprising angular distributions of the resulting products. The left-hand panel shows a two-dimensional representation of the angular distribution of fragments obtained from Sr_2 in a particular rotational quantum state, as predicted by quasiclassical theory; hot colours indicate higher distributions of fragments. The right-hand panel indicates the experimentally observed pattern, which can be explained only by using a full quantum-mechanical description of photodissociation.

Molecules can interact with light through either the light's oscillating electric field (which causes electric dipole transitions) or its oscillating magnetic field (magnetic dipole transitions). For most covalently bound molecules, the light intensity required to produce electric dipole transitions is a million times less than that required for magnetic dipole transitions. McDonald *et al.* are the first to have excited a pure magnetic transition and observed the fragments. This was possible because the Sr_2 molecules in this study are formed in the highest vibrational energy levels of the molecule's ground state, and therefore have a very long bond length, which increases the magnetic transition dipole moment by approximately 1,000-fold[9].

Another groundbreaking feature of McDonald and colleagues' work is that the Sr_2 molecules were prepared in a single rotational and vibrational quantum state by the laser-induced association of ultracold atoms, in the presence of an oriented magnetic field. Each state represents the projection (M) of a molecule's angular momentum vector (J) onto a quantization axis (in this case, the quantization axis aligns with the magnetic field). Several M states exist for each J value, and in the absence of a magnetic field they have the same energy (they are said to be degenerate); the number of M states is defined by the formula $2J+1$. When J is zero, it has no magnitude and no alignment in space. In a magnetic field, the M states do not have the same energy, because rotating electrons create a magnetic field that can be either aligned or counteraligned with the external magnetic-field quantization axis.

The authors' experiments started from a single (J,M) quantum state formed in the laser-association process. All of the quantum states reached during photodissociation were dictated by the starting state, and by the laser frequency and polarization relative to the magnetic field's axis. When the researchers obtained a single excited quantum state, they observed fragments recoiling predominantly parallel or perpendicular to the laser polarization. But if several degenerate quantum states were excited and interfered with each other, then the observed velocity distribution deviated spectacularly from purely parallel or perpendicular. These unexpected and previously unobserved angular distributions can be described only by a full quantum-mechanical treatment of the light-absorption process (Fig. 1).

At present, this sort of experiment is limited to a few diatomic molecules — some of which, like Sr_2, are not covalently bound — that can be generated by cold-atom techniques. However, there is much to be learnt from these studies, and as scientists learn to cool and trap a larger array of covalently bound molecules, the techniques developed and knowledge gained will provide the foundation for future research — for example, in polyatomic molecules. The photo-physics of polyatomic molecules is more complex than that of diatomic molecules, because multiple mechanisms couple their electronic states to each other, and several fragmentation pathways are possible. In the meantime, I personally found this article a joy to absorb. ■

© 2016 Macmillan Publishers Limited. All rights reserved.

David W. Chandler *is at the Combustion Research Facility, Sandia National Laboratories, Livermore, California 94550, USA.*
e-mail: chand@sandia.gov

1. McDonald, M. *et al. Nature* **535,** 122–126 (2016).
2. Busch, G. E., Mahoney, R. T., Morse, R. I. & Wilson, K. R. *J. Chem. Phys.* **51,** 449–450 (1969).
3. Solomon, J. *J. Chem. Phys.* **47,** 889–895 (1967).
4. Zare, R. N. & Herschbach, D. R. *Appl. Opt. (Suppl.)* **4,** 193–200 (1965).
5. Chandler, D. W. & Houston, P. L. *J. Chem. Phys.* **87,** 1445–1447 (1987).
6. Eppink, A. T. J. B. & Parker, D. H. *Rev. Sci. Instrum.* **68,** 3477–3484 (1997).
7. Heck, A. J. R. & Chandler, D. W. *Annu. Rev. Phys. Chem.* **46,** 335–372 (1995).
8. Ashfold, M. N. R. *et al. Phys. Chem. Chem. Phys.* **8,** 26–53 (2006).
9. McGuyer, B. H. *et al. Nature Phys.* **11,** 32–36 (2015).

CONSERVATION

The rainforest's 'do not disturb' signs

A study reveals that human-driven disturbances in previously undisturbed Amazon rainforest can cause biodiversity losses as severe as those of deforestation. Urgent policy interventions are needed to preserve forest quality. SEE LETTER P.144

DAVID P. EDWARDS

As we enter the Anthropocene, a proposed geological epoch shaped by human activity, mankind is driving a global biodiversity extinction crisis[1]. The conversion of forest to agricultural land is widely considered to be the leading cause of this crisis, especially in the hyperdiverse tropics[2], so avoiding deforestation is the predominant strategy for biodiversity conservation[3]. On page 144, Barlow *et al.*[4] present a landmark field study of Amazonian biodiversity in which they challenge the adequacy of this strategy by demonstrating the striking magnitude of several types of human-associated forest disturbance that are less immediately visible than deforestation.

Many studies have identified the negative effects on biodiversity of individual kinds of disturbance in tropical forests. These include the hunting of large animals[5], the selective logging of large, marketable trees[6], forest fires[7] and the creation of new edges to primary forests (those forests that have never been fully cleared) that, owing to deforestation, are buffeted by the hotter, drier and windier conditions found on adjacent farmland[8] (Fig. 1). However, by focusing on only one form of disturbance, such studies may have overlooked much greater conservation losses from the combined effects of forest disturbances.

Barlow and colleagues conducted biodiversity censuses across multiple landscapes and then developed a computational method for evaluating conservation losses (termed the 'conservation value deficit', a numerical value calculated by assessing biodiversity in disturbed primary forests relative to undisturbed ones). This enabled the authors to quantify the direct negative effects of deforestation and those resulting from the plethora of other types of forest disturbance.

The authors assembled an impressive data set collected across a large region of the Brazilian Amazon. They sampled 36 catchments (each 32–61 square kilometres in size) containing small rivers, spanning Belém and Tapajós, two major regions of endemism — areas that contain species that are found nowhere else. Each sample catchment varied in the degree of disturbance: 5 were entirely deforested, whereas the other 31 contained varying amounts of remnant forest, including undisturbed primary forests, and primary forests that had been disturbed by hunting, selective logging or fires, or isolated by surrounding farmland. Sampling the biodiversity across each catchment, Barlow and colleagues encountered a breathtaking total of 1,538 plant species, 460 bird species and 156 dung beetle species.

Their findings make for uncomfortable reading. Even catchments that retained 80% of their forest cover — the maximum that can be required of Amazonian estates under Brazil's Forest Code legislation — lost between 39% and 54% of their conservation value, and about half of this loss is due to disturbance within the remaining forest areas, rather than the losses from conversion to farmland. By extrapolating these disturbance-driven losses across the state of Pará, which represents 25% of the entire Brazilian Amazon, the authors found that conservation losses from disturbance are equivalent to the losses that would result from deforesting 92,000–139,000 km^2 of primary forest — an area roughly equivalent to the size of Greece.

They also found that species with higher conservation importance were more negatively affected by forest disturbance. Bird species that were restricted to small regions (with small global range sizes) fared worse than those with larger distributions, suggesting that forest disturbance is homogenizing biodiversity across regions[9]. Tree species with high wood density declined more than those with softer wood,

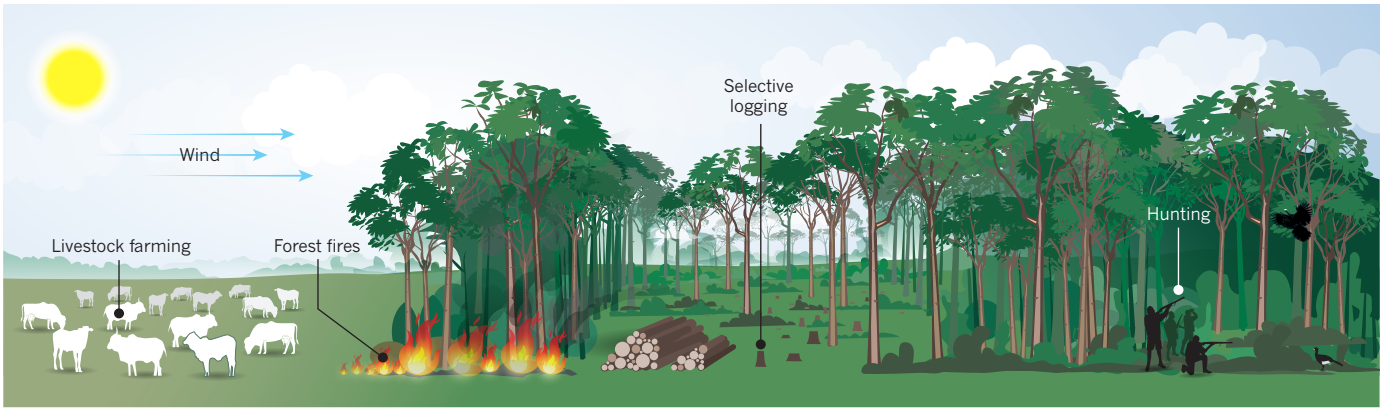


Figure 1 | Forest disturbance drives major conservation losses. Barlow *et al.*[4] report that the combined effects of various human-driven disturbances in the forests of the Brazilian Amazon can cause biodiversity losses on a scale similar to, or greater than, those caused by deforestation alone. Conversion to farmland can result in biodiversity loss and make forests more vulnerable to edge effects, such as the hot and windy conditions that can drive forest fires, which often ignite from farmland fires. Within the remaining rainforest, biodiversity can be affected by bushmeat hunting or selective logging.

© 2016 Macmillan Publishers Limited. All rights reserved.

degrading the ability of disturbed forests to store carbon, thereby driving climate change[10].

Barlow and colleagues' study has three limitations worthy of comment, each of which probably means that their assessment of biodiversity loss from disturbance is conservative. First, the extensive human activity in the study regions means that many of their undisturbed forest plots could have suffered from low-level or historical disturbances that were not detected. Although the authors attempted to correct for this potential bias in their analytical approach, if they had sampled truly remote and undisturbed forests instead, densities of the most sensitive species would probably have been higher, and the conservation value deficit would thus have been even more severe in disturbed forests.

Second, the authors' results from two areas with species endemism were extrapolated to a quarter of the Brazilian Amazon, which spans an additional three unstudied areas of endemism. Their analysis therefore fails to fully capture the species differences between regions[9]. Small-ranged species will probably be more negatively affected than those with larger distributions.

Finally, mammals were not studied, yet they have crucial roles in maintaining healthy ecosystems[10]. Mammals would be expected to suffer at least as profoundly as the sampled taxa, owing to the severity of hunting in accessible forests located among farmlands or near roads[11]. More generally, whether findings from the Brazilian Amazon can be extrapolated to other tropical regions is unknown. Barlow *et al.* have given scientists the impetus and methodological tools to make such assessments, making similar studies a research frontier elsewhere in the tropics.

This research challenges the governance of tropical forests, and hence the protection of the myriad conservation and ecosystem benefits they provide, which sustain some of the most biodiverse ecosystems on the planet and the livelihoods of millions of people. As the authors acknowledge, avoiding deforestation must remain a key tenet of conservation strategies[3]. However, their results underscore the need for a step-change in forest governance, with much greater emphasis on the ecological health of retained forest[1]. In particular, Barlow and colleagues have shown that policies, such as Brazil's Forest Code, that set targets for forest cover without also setting requirements for forest quality are insufficient to prevent substantial conservation losses, and are a slippery slope to greatly impoverished ecosystems[10].

The authors' results underscore the need for a step-change in forest governance.

To remedy this, agricultural landscapes must be better designed to promote the protection of larger and less-isolated forest blocks. More-stringent regulation and enforcement are needed, both of fire use in agriculture (which frequently spills into forests) and of selective logging — together with clear economic benefits for more sustainably managed agriculture and sustainable logging[6]. These all require coordination between landowners, policymakers and conservationists across entire landscapes and regions.

In many tropical forest regions, disturbed primary forests that have seen big biodiversity losses are especially valuable for conservation, because large undisturbed forests are rare or completely lacking. Within such regions, these results underscore the necessity of assisting the recovery of disturbed forests or of taking unproductive farmland out of use to restore forest coverage and connectivity. Although the biodiversity extinction crisis could be even worse than currently recognized, by embracing better management strategies, the solutions are still within our grasp. ■

David P. Edwards *is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.*
e-mail: david.edwards@sheffield.ac.uk

1. Lewis, S. L., Edwards, D. P. & Galbraith, D. *Science* **349,** 827–832 (2015).
2. Newbold, T. *et al. Nature* **520,** 45–50 (2015).
3. Nepstad, D. *et al. Science* **326,** 1350–1351 (2009).
4. Barlow, J. *et al. Nature* **535,** 144–147 (2016).
5. Peres, C. A. *Conserv. Biol.* **14,** 240–253 (2000).
6. Edwards, D. P., Tobias, J. A., Sheil, D., Meijaard, E. & Laurance, W. F. *Trends Ecol. Evol.* **29,** 511–520 (2014).
7. Barlow, J. & Peres, C. A. *Ecol. Appl.* **14,** 1358–1373 (2004).
8. Ferraz, G. *et al. Science* **315,** 238–241 (2007).
9. Socolar, J. B., Gilroy, J. J., Kunin, W. E. & Edwards, D. P. *Trends Ecol. Evol.* **31,** 67–80 (2016).
10. Peres, C. A., Emilio, T., Schietti, J., Desmoulière, S. J. M. & Levi, T. *Proc. Natl Acad. Sci. USA* **113,** 892–897 (2016).
11. Laurance, W. F., Goosem, M. & Laurance, S. G. W. *Trends Ecol. Evol.* **24,** 659–669 (2009).

This article was published online on 29 June 2016.

COMPUTATIONAL NEUROSCIENCE

Species-specific motion detectors

A range of neuronal mechanisms can enable animals to detect the direction of visual motion. Computational models now indicate that a factor as simple as eye size might explain some of this diversity. SEE ARTICLE P.105

THOMAS EULER & TOM BADEN

Seeing whether and where an object moves is crucial for the survival of any visually oriented animal, whether predator or prey. Consequently, motion and its direction are computed at many levels along the vertebrate visual pathway, starting in the retina. One key element of direction-selective retinal neuronal circuits is the starburst amacrine cell (SAC). On page 105, Ding *et al.*[1] unpick the mechanisms that mediate SAC direction selectivity in the mouse retina.

Structures called dendrites project radially from the body of the SAC to give the cell its characteristic shape, which resembles an exploding star. The dendrites receive excitatory inputs from the retina's light-sensing photoreceptor cells via bipolar cells, and in turn make inhibitory synaptic connections to neurons called direction-selective ganglion cells (DSGCs) and other SACs. Different types of DSGC are each robustly tuned to movement in a particular (preferred) direction.

The work of numerous labs over past decades has indicated that the inhibitory signals sent by SACs to DSGCs contain information about movement direction. However, an SAC as a whole is not selective for one particular direction; instead, each dendrite is tuned to the direction of movement that is aligned with the direction from the cell body to the dendritic tip[2]. In addition, SAC dendrites tuned to a particular direction make more synaptic connections with DSGCs that prefer the opposite direction than with those of the same preference, providing DSGCs with information that defines their tuning[3].

Although this general layout is broadly accepted, the mechanism that renders SAC synaptic outputs direction-selective is still intensely debated. In mammals, several (not necessarily mutually exclusive) mechanisms have been proposed. Some rely on properties of the dendrites — for instance, the spatial arrangement of channel proteins in the membrane or of a chloride gradient along the dendrite. Others invoke network interactions such as reciprocal inhibition between SACs or a particular spatial arrangement of bipolar-cell inputs that signal to the neuron with different timings (reviewed in refs 4 and 5). But the relative contribution of each mechanism is unclear.

© 2016 Macmillan Publishers Limited. All rights reserved.

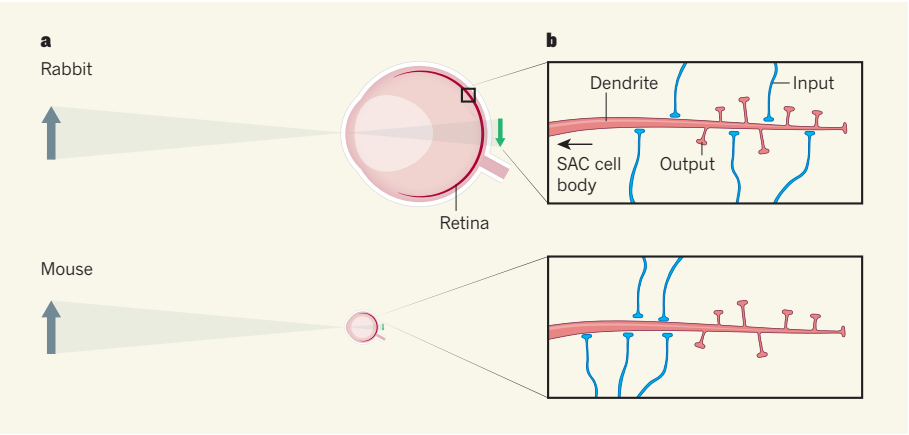


Figure 1 | Direction is in the eye of the beholder. **a**, Rabbits have larger eyes than mice. Therefore, the image of an object travelling a given distance (grey arrows) will traverse different distances across the retina of each species (green arrows). Ding *et al.*[1] hypothesize that neuronal circuits in mice thus need to respond to lower retinal-image velocities to compute information about movement direction. **b**, The authors find that the synaptic connections to direction-selective starburst amacrine cells (SACs) in the retina differ between the two species. As shown in this simplified schematic, inputs (both from other SACs and from bipolar cells) cover the length of the cells' radially projecting dendrites in rabbits, whereas inputs and outputs are well segregated along the dendrites in mice. This difference increases the directional tuning of mouse SAC dendrites at slower retinal velocities.

Ding *et al.* investigated the role of network interactions in SAC direction selectivity in mice. They used a large electron-microscopy data set to generate a map of all the synaptic connections to and from SACs. This allowed them to precisely assess the spatial arrangement of both input and output synapses along SAC dendrites. Synapses followed a clear pattern: inputs (both excitatory and inhibitory) were restricted to the proximal section of the dendrites, whereas output synapses were located at the dendritic tips (Fig. 1).

Although this output arrangement comes as no surprise, the proximal restriction of inhibitory inputs is in stark contrast to the situation seen in rabbit SACs, which show much less spatial segregation[6]. This is puzzling. Why should two ground-dwelling mammals that live in similar environments use different solutions to compute motion direction?

To address this question, Ding and colleagues used a computer model to simulate how different synaptic arrangements might affect SAC direction selectivity. They connected a single SAC to bipolar cells and neighbouring SACs. They then varied the arrangement of input synapses along the central SAC's dendrites to mimic the mouse and rabbit 'solutions', and compared the cells' performance. The higher degree of segregation in mice generated more-robust directional tuning, in particular for stimuli that traversed the retinal surface slowly. The authors then went on to confirm these predictions experimentally by recording motion-evoked signals in mouse SAC dendrites.

Why would mice need to compute slower movements than rabbits? After all, the velocities of movement that these animals encounter in the wild are probably similar. The answer is deceptively simple. Mice have smaller eyes than rabbits. A moving object's angular velocity (velocity measured as the angle travelled per unit of time instead of distance per unit of time) translates to a lower absolute velocity across the retinal surface of a smaller eye than that for a larger eye. Therefore, the object's image moves substantially more slowly on the mouse's retinal surface (Fig. 1). Perhaps the less-segregated synaptic arrangement in the rabbit circuit is good enough to encode the relevant range of stimulus velocities, whereas mice needed to evolve a solution that also reliably works for slower movements across the retina.

This is a neat study for several reasons. First, it represents a well-balanced synthesis of large-scale, high-resolution circuit anatomy, realistic modelling and synaptic neurophysiology. Second, demonstrating that a simple difference such as eye size can have a direct impact on how circuits and computations are implemented highlights the often-underestimated importance of taking species differences into account. The suggestion that different species may use different adaptations of a general computation for retinal direction selectivity could be the key to reconciling seemingly contradictory findings in the field.

Third, this study takes a crucial step towards the development of a truly integrated model of direction-selective retinal circuitry. A carefully extended version of the model designed by Ding and co-workers (for instance, one that includes more-realistic bipolar inputs) could be instrumental in disentangling different direction-selectivity mechanisms. In addition, this approach will allow researchers to systematically address other mysterious aspects of retinal direction detection, such as the role of the molecule acetylcholine, which is SACs' secondary neurotransmitter and seems to play a part in signalling only under certain stimulus conditions[7,8].

Retinal direction-selective circuits should now be studied in other species, perhaps starting with those that have extreme eye sizes or more-distant evolutionary roots. For instance, the DSGCs of zebrafish larvae have largely similar properties to those of mammalian DSGCs[9], suggesting a similar organization of direction-selective circuits. However, in the tiny larval eye, an object moving at an angular velocity of 1 degree per second crosses the retinal surface at a mere 3 micrometres per second — 10 times more slowly than in mice. Maybe zebrafish have an even more precise synaptic arrangement along SAC dendrites. Or, perhaps more likely, they have another direction-detection mechanism altogether.

Primate and rabbit eyes are not that different in size. However, direction selectivity in the primate retina is a puzzle. SACs are present, but are sparser than in any other mammal studied[10]. Despite long-standing and vigorous attempts, there is no direct evidence so far that primates have DSGCs (discussed in ref. 11). Instead, primates seem to generate direction-selective responses farther along the visual pathway.

What is the take-home message? Maybe, that the goal of neuroscience is not to 'solve' the mouse, rabbit or zebrafish. Instead, neuroscientists should collect different solutions to common computational problems. Which solutions are actually implemented in any particular instance is perhaps secondary. After all, neuroscience is about building an understanding of the general principles by which neurons and networks generate function. ■

Thomas Euler *and* **Tom Baden** *are at the Institute for Ophthalmic Research, University of Tübingen, 72076 Tübingen, Germany.*
T.E. *is also in the Center for Integrative Neuroscience, University of Tübingen.*
T.B. *is also at the School of Life Sciences, University of Sussex, Brighton, UK.*
e-mails: thomas@eulerlab.de; tom@badenlab.org

1. Ding, H., Smith, R. G., Poleg-Polsky, A., Diamond, J. S. & Briggman, K. L. *Nature* **535,** 105–110 (2016).
2. Euler, T., Detwiler, P. B. & Denk, W. *Nature* **418,** 845–852 (2002).
3. Briggman, K. L., Helmstaedter, M. & Denk, W. *Nature* **471,** 183–188 (2011).
4. Vaney, D. I., Sivyer, B. & Taylor, W. R. *Nature Rev. Neurosci.* **13,** 194–208 (2012).
5. Kim, J. S. *et al. Nature* **509,** 331–336 (2014).
6. Famiglietti, E. V. *J. Comp. Neurol.* **309,** 40–70 (1991).
7. Grzywacz, N. M. & Amthor, F. R. *Vis. Neurosci.* **24,** 647–661 (2007).
8. Lee, S., Kim, K. & Zhou, Z. J. *Neuron* **68,** 1159–1172 (2010).
9. Gebhardt, C., Baier, H. & Del Bene, F. *Front. Neural Circuits* **7,** 111 (2013).
10. Rodieck, R. W. *J. Comp. Neurol.* **285,** 18–37 (1989).
11. Borst, A. & Euler, T. *Neuron* **71,** 974–994 (2011).

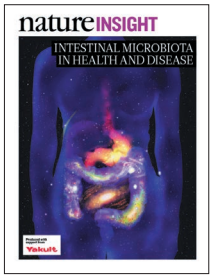
This article was published online on 22 June 2016.

© 2016 Macmillan Publishers Limited. All rights reserved.

nature INSIGHT

INTESTINAL MICROBIOTA IN HEALTH AND DISEASE

7 July 2016 / Vol 535 / Issue No 7610



Cover illustration
Jessica Fortner

Editor, *Nature*
Philip Campbell

Publishing
Richard Hughes

Insights Editor
Ursula Weiss

Production Editor
Elizabeth Batty

Art Editor
Nik Spencer

Sponsorship
Yuki Fujiwara
Yvette Smith

Production
Ian Pope

Marketing
Nicole Jackson
Alan Abery
Aiko Shuzui

Editorial Assistant
Giacomo Russo

The Campus
4 Crinan Street
London N1 9XW, UK
Tel: +44 (0) 20 7833 4000
e: nature@nature.com

SPRINGER NATURE

The human gut is home to trillions of microorganisms, which modulate health and disease. This Insight brings together leaders in the field of microbiota–host interactions to provide an overview of basic biological processes and important advances in the development of clinical applications.

Jeff Gordon and colleagues present a microbial perspective of human developmental biology. They describe how the microbiota affects prenatal and postnatal growth and explain how an understanding of such communities could help to prevent and treat diseases. To this end, they call for the establishment of 'human microbial observatories' to examine the development of the microbiota in birth cohorts with diverse lifestyles and patterns of disease.

Justin Sonnenburg and Fredrik Bäckhed analyse how the microbiota and diet interact to influence metabolism. They review mechanisms used by the microbiota to modulate the effects of diet on the host's metabolic status, as well as the potential for therapeutic intervention.

Eran Elinav and colleagues discuss crosstalk between the microbiota and the innate immune system, focusing on bacterial components and host response pathways, mutually beneficial effects of such communication and diseases that arise when this interaction is disturbed.

Kenya Honda and Dan Littman summarize our understanding of how specific microbes determine aspects of adaptive immunity and the part that they play in the induction of both immune tolerance and conditions such as allergy and intestinal inflammation.

Andreas Bäumler and Vanessa Sperandio examine interactions between the gut microbiota and pathogenic bacteria, including how pathogenic species exploit microbiota-derived sources of carbon and nitrogen as nutrients and regulatory signals for growth and virulence.

And Rob Knight and colleagues consider the advent of microbiome-wide association studies, which have been enabled by advances in DNA sequencing, metabolomics, proteomics and computation. They provide a road map for realizing the promise of microbiome-based precision diagnostics and therapies.

Nature is pleased to acknowledge the financial support of Yakult Honsha Co., Ltd in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content.

Christina Tobin Kåhrström, Nonia Pariente & Ursula Weiss
Senior Editors

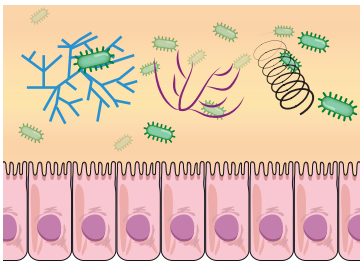
CONTENTS

PERSPECTIVE

48 **A microbial perspective of human developmental biology**
Mark R. Charbonneau, Laura V. Blanton, Daniel B. DiGiulio, David A. Relman, Carlito B. Lebrilla, David A. Mills & Jeffrey I. Gordon

REVIEWS

56 **Diet–microbiota interactions as moderators of human metabolism**
Justin L. Sonnenburg & Fredrik Bäckhed

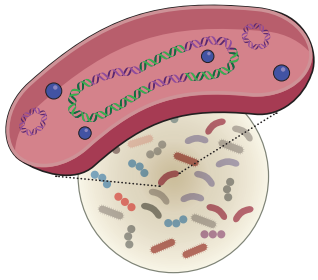


65 **The microbiome and innate immunity**
Christoph A. Thaiss, Niv Zmora, Maayan Levy & Eran Elinav

75 **The microbiota in adaptive immune homeostasis and disease**
Kenya Honda & Dan R. Littman

85 **Interactions between the microbiota and pathogenic bacteria in the gut**
Andreas J. Bäumler & Vanessa Sperandio

94 **Microbiome-wide association studies link dynamic microbial consortia to disease**
Jack A. Gilbert, Robert A. Quinn, Justine Debelius, Zhenjiang Z. Xu, James Morton, Neha Garg, Janet K. Jansson, Pieter C. Dorrestein & Rob Knight



© 2016 Macmillan Publishers Limited. All rights reserved.

PERSPECTIVE

doi:10.1038/nature18845

A microbial perspective of human developmental biology

Mark R. Charbonneau[1,2], Laura V. Blanton[1,2], Daniel B. DiGiulio[3,4], David A. Relman[3,5], Carlito B. Lebrilla[6,7], David A. Mills[7,8,9] & Jeffrey I. Gordon[1,2]

When most people think of human development, they tend to consider only human cells and organs. Yet there is another facet that involves human–associated microbial communities. A microbial perspective of human development provides opportunities to refine our definitions of healthy prenatal and postnatal growth and to develop innovative strategies for disease prevention and treatment. Given the dramatic changes in lifestyles and disease patterns that are occurring with globalization, we issue a call for the establishment of 'human microbial observatories' designed to examine microbial community development in birth cohorts representing populations with diverse anthropological characteristics, including those undergoing rapid change.

A survey of the biological landscape that encompasses human development should consider all facets of what it means to be 'human'. There are at least as many microbial cells as there are human cells in our bodies, and the vast majority of unique genes are microbial[1–3]. As such, we can view ourselves as holobionts[4]. The dynamic microbe–microbe and microbe–host interactions that allow our microbial communities to assemble and endure are as yet largely uncharacterized. Our relationships with microbes begin before birth; they represent potentially modifiable features of postnatal development, and probably contribute to intra- and interpersonal variations in many aspects of normal physiology, metabolism, immunity and neurology, as well as to predisposition to diseases.

The past decade has produced a magnificent and still rapidly evolving toolbox of experimental and computational techniques for culture-independent identification of the microorganisms that comprise our body habitat-associated microbial communities (microbiota), as well as their genes (microbiome) and gene products. These tools allow a number of hypotheses about microbial contributions to human development to be tested. One hypothesis is that maternal microbial ecology affects pregnancy, fetal development and the future health of offspring. If true, the hypothesis suggests the possibility of prenatal prognostic and diagnostic measurements and therapeutic interventions that target the maternal microbiota to guide healthy fetal development and avoid premature birth and other negative outcomes. Another hypothesis is that after birth, there are microbial taxa whose changing patterns of representation can be used to define 'normal' programmes of development of the microbial communities that occupy a given body habitat in biologically unrelated individuals with healthy growth phenotypes (as defined by anthropometric indices). A corollary to this hypothesis is that deviations from these normal programmes of community assembly represent a way to characterize abnormal development, including states of immaturity or precocious maturation. Establishing a causal relationship between the state of microbial community development and healthy growth would allow deviations from normal microbiota development to be used as a parameter for risk assessment or classification of a number of diseases that may manifest themselves early or later in life, yield insights about disease pathogenesis, and provide a starting point for developing microbiota-directed therapeutic interventions or new approaches for disease prevention.

In this Perspective, we discuss evolving concepts about the relationship between maternal microbial ecology (before, during and after pregnancy) and pregnancy outcomes as well as the relationship between human breast milk oligosaccharides, the establishment and expressed functions of the gut microbiota and healthy postnatal growth. We also address the need for long-term birth cohort studies to identify both shared and distinctive features of microbial community development, within and across populations, and delineate how normal execution (and perturbations) of this facet of human developmental biology is related to health status.

Maternal microbial ecology

The structure and function of maternal microbial communities, and the impact of these communities on maternal and infant health outcomes has been considered in several body habitats, including the vagina, the distal gut and the mouth.

Vaginal microbiota

For decades, culture-based studies have suggested that lactobacilli are the most prevalent constituents of the vaginal microbiota in non-pregnant and pregnant women[5]. More recently, culture-independent studies have demonstrated that most vaginal communities are dominated numerically by a single *Lactobacillus* species. This finding has prompted some investigators to assign vaginal communities to a relatively limited number of discrete 'community state types' (CSTs). These CSTs are classified either by which *Lactobacillus* species is dominant (CST I, II, III and V) or by the presence of a relatively diverse, *Lactobacillus*-poor community (CST IV)[6]. The resolution and veracity of the vaginal CST model remains unsettled: some investigators have proposed other stable or transitional states beyond the five described initially[7]. Others have highlighted potential pitfalls, including the extent to which the detection of state types is dependent on the analytical workflow[8]. Irrespective of the ultimate usefulness of the CST model, the limited diversity

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, Missouri 63110, USA. [2]Center for Gut Microbiome and Nutrition Research, Washington University School of Medicine, St Louis, Missouri 63110, USA. [3]Department of Medicine, Stanford University, Stanford, California 94305, USA. [4]VA Palo Alto Health Care System, Palo Alto, California 94304, USA. [5]Department of Microbiology and Immunology, Stanford University, Stanford, California 94305, USA. [6]Department of Chemistry, University of California, Davis, Davis, California 95616, USA. [7]Foods for Health Institute, University of California, Davis, Davis, California 95616, USA. [8]Department of Food Science and Technology, University of California at Davis, Davis, California 95616, USA. [9]Department of Viticulture and Enology, University of California, Davis, Davis, California 95616, USA.

© 2016 Macmillan Publishers Limited. All rights reserved.

of abundant taxa in vaginal communities suggests that a deterministic process of community assembly, such as habitat filtering, governs the overall structure of the adult vaginal microbiota.

CST IV is similar to the microbiota structure encountered in bacterial vaginosis, a dysbiosis that is associated with adverse health outcomes, including preterm birth[9,10]. In non-pregnant North American women, the prevalence of CSTs varies with self-reported race and ethnicity. CST IV is observed in about 40% of African American and Hispanic women, but only about 20% of Asian American women and about 10% of Caucasian women[6]. This skewed distribution suggests that a diverse, non-*Lactobacillus*-dominated community (CST IV) might represent a normal variant in a subset of women and argues for an expanded assessment of what comprises a healthy vaginal microbiota.

Little is known about the development of the vaginal microbiota before and after puberty, or how different vaginal community 'fates' (structural and functional states) in adulthood are determined. One area that should be investigated is the relationship between the glycan content of the vaginal mucosa and the community state, including the biogeographical features of each state. In addition, much remains to be learned about the effects of bacterial and eukaryotic taxa (and the viruses they host) on vaginal epithelial-cell differentiation, vaginal mucosal metabolism and the activities of components of the innate and adaptive arms of the immune system that are represented in this habitat. The development of microarrays composed of purified microbial glycans[11] provides one way to characterize immunological responses to the bacterial antigens represented in the vaginal microbiota and thus creates another approach to classify community states. Representative preclinical models are needed for testing whether causal relationships exist between these and other environmental factors and community states. They will also help to characterize the mechanisms that shape community assembly, that determine community responses to various perturbations, that underlie community resiliency and that mediate the effects of community states on host biology.

A compelling question is whether there is a discernable programme of change in the properties of the vaginal microbiota before, during and after pregnancy and, if there is, the extent to which such change recapitulates features of the original developmental biology of the community. A related question is whether and how functional alterations in vaginal microbial community states and in the microbiota at other body sites during pregnancy affect intrauterine growth of the fetus (Box 1). Work has focused on bacterial taxonomic composition of these communities rather than on the functional features they express. Studies currently suggest that the bacterial composition of the microbiota is more stable during pregnancy than at other times during adulthood[12–15]. The diverse CST IV seems to be the least stable community state during pregnancy: it exhibits a substantially higher rate of transition to alternative CSTs on a week-to-week timescale than do the four *Lactobacillus*-dominated CSTs[14]. A note of caution is that vaginal microbial community composition has not been defined in time-series studies in which samples are taken from the same women before conception, during and after pregnancy, and during subsequent pregnancies. In addition, little is known about the non-bacterial membership of the pregnancy-associated microbiota.

Some factors are thought to promote vaginal microbiota structural stability during pregnancy, such as a lack of menses. However, many factors remain unknown, as is the degree to which structural stability is accompanied by functional stability and how this relates to the transfer of taxa from mothers to infants during the immediate postpartum period. From an anthropological perspective, it is interesting to note that prescribed diets during pregnancy are an important part of the cultural traditions of some populations[16]. It is unclear how these treatments affect vaginal (and gut) microbial community structure, function and stability. The answers to these questions could yield fresh approaches for deliberate manipulation of the vaginal microbiota.

In contrast to the structural stability seen during pregnancy, studies of women in the United States, Europe, Africa and Asia have shown that, after delivery, the vaginal microbiota commonly undergoes an abrupt and striking alteration in its taxonomic composition[14,17–19]. This alteration is characterized by a significant increase in within-community (or α) diversity and is driven by a decrease in the abundance of *Lactobacillus* species and a commensurate increase in a wide range of anaerobic species. Although many features of altered postpartum microbial communities remain to be elucidated (such as the time it takes to return to the 'baseline' state), it seems that they can persist for at least 1 year in many women[14]. A short interval (less than 12 months) between pregnancies is associated with an increased risk of preterm birth; whether a persisting altered postpartum vaginal community contributes to this risk warrants further study.

BOX 1

The evolving maternal–fetal microbial landscape

Much remains to be learned about the assembly, host interactions and transmission of maternal and early childhood microbial communities. Hypotheses about the evolving maternal–fetal microbial landscape deserve further attention and testing.

- Some activities of the maternal microbiota might have a beneficial impact on fetal nutrition and development.
- Altered compositions and expressed activities of the maternal microbiota might contribute to gestational outcomes, including adverse events such as premature labour and birth.
- Microbes that are transferred to offspring before or during delivery might reflect environmental exposures of the mother during pregnancy (for example, diet).
- Persisting disturbances in the vaginal microbiota after giving birth might pose a risk for preterm delivery in subsequent pregnancies.
- Variations in the transfer of microbes from mothers to infants might affect early postnatal development of the child's microbiota, immune system and metabolic processes.

Gut microbiota

Much more information is needed about whether the structural and functional properties of the gut microbiota of women change as a function of pregnancy. If changes do occur, whether and how they relate to maternal and fetal health, as well as the subsequent health of infants and children, should be investigated. The relationship between maternal nutritional status at the time of conception and the health of the newborn is well established[20]. A study of pregnant Finnish women reported a significant increase in faecal energy content, as determined by bomb calorimetry, between the first and third trimesters despite stable diets and energy intake[21]. This change in energy content correlated with shifts in taxonomic composition[21]. However, studies of women residing in the United States[14] and in Tanzania[17], which were conducted at higher temporal resolution, found that the women's faecal microbiota manifested compositional stability throughout pregnancy (as measured by trends of α diversity, week-to-week variation in bacterial composition within subjects and β diversity across gestational time). The reasons for these divergent findings are unclear. The maternal microbiota and diet also have the potential to influence both fetal and maternal epigenomes, although a discussion of this topic is beyond the scope of this Perspective.

Oral microbiota

Mothers harbour complex microbial communities in their mouths. The composition[22] and transcriptional activities[23] of these communities are altered in the setting of periodontitis, a condition also associated with

© 2016 Macmillan Publishers Limited. All rights reserved.

intrauterine growth restriction, preterm birth and low birthweight[24]. A study of the oral microbiota of women living in the United States and Africa indicated that the taxonomic composition remains stable during pregnancy[14,17]. However, pre-conception data from the same women were unavailable for comparison. Microbial taxa have been detected in amniotic fluid[25–28] and in the placenta[29] that probably originate from the mouth, particularly in women who are unhealthy or had adverse outcomes such as preterm labour with intact fetal (chorioamniotic) membranes or premature rupture of these membranes. Disentangling adverse effects on pregnancy that originate from the oral microbiota is challenging, especially if disease results from a perturbation in relatively minor constituents of the community[30].

Development of the oral microbiota has not been comprehensively defined through time-series studies of healthy infants and children. For example, the effects of maternal prenatal history, gestational age, route of delivery and milk-feeding history remain to be characterized. One study attempted to define 'normal' development by following 50 children from the ages of 4 years to 6 years[31]. A strong effect of chronological age was observed on the taxonomic composition of the oral microbiota. This effect was more pronounced for bacterial communities in supragingival plaque than in saliva, which suggests body-habitat-specific differences in community assembly programmes. Deviations from early, normal community compositions were predictive of subsequent development of dental caries[31].

Preterm delivery and fetal exposure to microbes

The extent to which the fetal environment is sterile has been pondered since the birth of the field of microbiology[32]. Early studies suggested that the amniotic cavity was universally sterile before labour[33], although subsequent, indirect evidence has challenged that assumption[34]. Culture-based and later, polymerase-chain-reaction-based studies indicated that microbial invasion of the amniotic cavity occurs more frequently and involves a greater diversity of microbes than was originally thought[25,26]. Endometrial sampling of the intrauterine cavity in non-pregnant women has yielded widely varying rates (0–89%) of microbial recovery across culture-based studies[35]. Molecular-based studies suggest that most uteruses harbour microbes, with *Lactobacillus*, *Prevotella* and *Bacteroides* among the genera that are most commonly encountered[35,36]. However, data obtained during pregnancy are lacking.

At the time of delivery, the basal plate of the placenta contains intracellular bacteria in about a quarter of women, but in about half of those who deliver spontaneously before 28 weeks of pregnancy[37]. One study has shown that the placenta harbours a complex set of microbial DNA sequences[38]. But unlike more densely colonized body sites such as the gut and mouth, placental samples are overwhelmingly negative in culture-based assays[39]. DNA-based assessments of potential microbes in the placenta, and other low microbial biomass sites, are particularly prone to confounding findings from 'background' DNA[40,41] and should be interpreted with caution in the absence of appropriate controls. The degree to which the fetal–placental environment has evolved to serve as a venue for programmed engagement of diverse microbes, as opposed to being a site that simply tolerates stochastic low-level microbial exposures, remains unclear and merits further study.

A report published this year suggests that in women who experienced spontaneous preterm birth, those with histological evidence of severe chorioamnionitis have fewer species of bacteria on the fetal side of the placental membrane than do those who do not have severe chorioamnionitis[42]. This difference might be driven by a high abundance of a limited number of clonal pathogens (as is typical of many clinical infections) in women with severe chorioamnionitis. Further studies with appropriate negative controls are needed to corroborate these findings and to resolve unanswered questions such as the body site of origin of the detected microbes, as well as the direction and timing of their translocation across adjacent tissues[43].

Microbes have been detected in first-pass meconium samples from approximately two-thirds of healthy, vaginally delivered, breastfed full-term babies, but at very low levels[44]. Detection is more common in meconium from neonates who are born before 33 weeks of gestation, and there is considerable taxonomic overlap with the microbes found in the amniotic fluid[25,45]. Molecular evidence for microbial invasion of the amniotic cavity has provided associations of space, time and 'dose' that support a causal relationship with preterm birth[25]. Microbial taxa that are associated with preterm birth most frequently originate from the mother and exploit one of three natural routes for invading the amniotic cavity[46]: ascension from the vagina and cervix; transfer through the fallopian tubes; or translocation from more distant sites of colonization in the body, presumably through the bloodstream[27]. The majority of invading microbes seem to come from the vagina[25,28,46], although other body habitats, most notably the mouth[27,47] and gut[26], may have a role in some cases. Taxa associated with CST IV communities, such as *Ureaplasma* and *Prevotella* species, are among the more common invaders. By contrast, *Lactobacillus* species are rarely encountered in amniotic fluid, even after membrane rupture[26]. This suggests that features of specific microbial taxa, or groups of taxa that occur together in CST IV communities, underpin factors that promote invasion of the amniotic cavity, such as virulence genes and divergent host immune responses[48]. Whether particular vaginal CSTs or the presence and abundance of individual taxa are associated with preterm birth is an unresolved question of great interest. Studies have produced conflicting results[12,14]. If vaginal CST IV communities are indeed associated with preterm birth in some women, this would be broadly consistent with epidemiological evidence that links bacterial vaginosis, which shares taxonomic similarity with CST IV communities, to an increased risk of preterm birth[9].

The effect of preterm delivery on the development of microbial communities in premature babies has been examined mainly from the perspective of the infant. Comparing the development of microbial communities in premature and full-term infants could lead to amended or new definitions of biological immaturity. Such definitions are confounded, however, by the frequent pre-emptive administration of antibiotics to babies who are born prematurely. Maternal microbial communities may also exert a significant influence. An elegant study demonstrated that transient microbial colonization of pregnant, germ-free mice was sufficient to modulate the function of innate immune cells in the small intestines of their germ-free offspring[49]. Microbial products were detected in both the dam's milk and placenta, which suggests that 'indirect' exposure to microbes through the mother is sufficient to shape neonatal development. Such findings suggest that systematic characterization of multiple body-habitat-associated microbial communities in mothers who have preterm versus full-term pregnancies creates opportunities to examine whether there are identifiable programmes of change in maternal microbial ecology during pregnancy and whether disruption of these programmes affects initial transfer of microbes to their offspring (and the subsequent development of the children's microbiota). This knowledge could change clinical practice so that more attention is placed on careful stewardship of microbial resources in women who have a high risk of preterm delivery[50]. Deliberate efforts could be made to transfer these microbes to their offspring, with the potential for supplementation with important taxa that are missing.

Breast milk and infant gut microbiota

Researchers are beginning to uncover how breast milk composition changes over time after parturition and how it shapes the structural and functional maturation of infant-associated microbial communities.

Microbes associated with breast milk

Studies of milk-associated microbiota reveal highly individualized assemblages[51]. These groups of microbes are routinely dominated by skin-associated bacteria, such as staphylococci and streptococci, which generally do not persist in the infant gut in significant numbers for more than a few weeks[52]. Some anaerobic species, such as *Bifidobacterium*, have been isolated from breast milk, which suggests a route for transit of specific strains that eventually colonize the infant colon. The factors

© 2016 Macmillan Publishers Limited. All rights reserved.

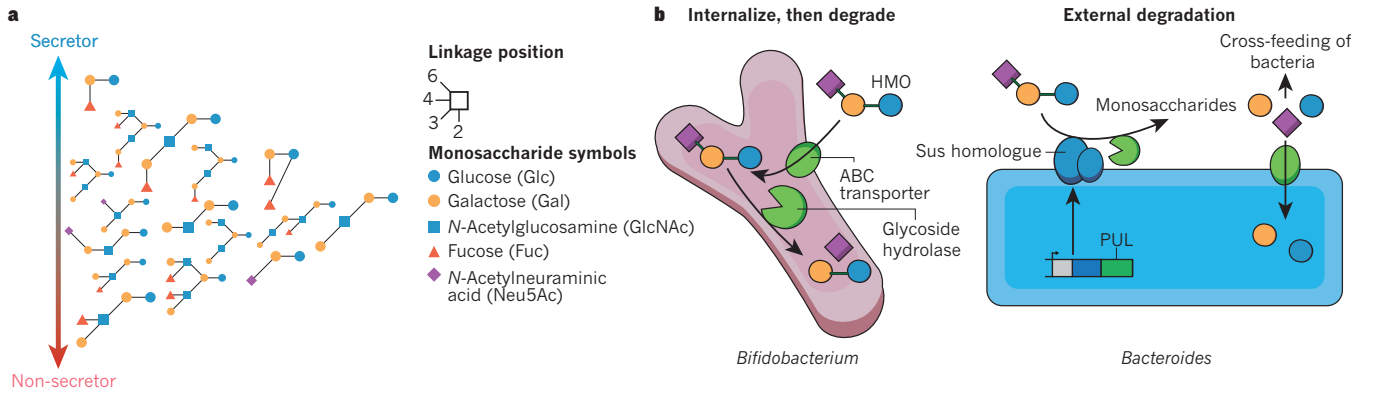


Figure 1 | Oligosaccharides in human breast milk and strategies for their degradation by the infant microbiota. a, HMOs that are most abundant in the breast milk of mothers who are secretors are indicated by the blue arrow; those that are most abundant in the breast milk of non-secretors are indicated by the red arrow. Structures at the intersection of the arrows are found in both secretor and non-secretor mothers in similar abundances. Monosaccharides in HMOs, as well as their glycosidic linkages, are described by the inset key. **b,** Most strains of *Bifidobacterium* (left) use an 'internalize, then degrade' strategy in which HMO structures are first imported using ABC transporters and then degraded by intracellular glycoside hydrolases. Strains of *Bacteroides* (right) typically employ an 'external degradation' strategy for HMO structures, which involves cell-surface-associated carbohydrate-binding proteins and secreted glycoside hydrolases that are encoded by polysaccharide utilization loci (PULs). These PULs have features similar the prototypic starch utilization system (Sus) of *Bacteroides thetaiotaomicron*. This external degradation can result in 'cross-feeding' of secondary consumers, including potentially pathogenic bacteria, in the infant gut microbiota.

that contribute to the strain-specific composition in breast milk are unclear and are subject to debate.

Human milk oligosaccharides

From a molecular perspective, breast milk is the best-characterized food that humans consume. The most abundant component in dried samples of breast milk is lactose, which provides nutrition for the infant, although many bacterial taxa can also digest this disaccharide. Lactose is made available specifically to bacterial colonizers of the infant gut by extending it by 3–20 monosaccharide units to yield structures that are known collectively as human milk oligosaccharides (HMOs)[53,54]. All HMOs contain this lactose core together with various combinations of glucose, galactose, *N*-acetyl galactosamine, fucose and sialic acid (*N*-acetylneuraminic acid or Neu5Ac)[55]. HMOs are often terminated by fucose or sialic acid (Fig. 1a). Approximately 60% of HMO structures are fucosylated, and 5–20% are sialylated[56,57].

The role of HMOs has become more apparent through the application of nanoflow liquid chromatography mass spectrometry. This method has detected more than 300 HMO structures in breast milk samples pooled from several mothers, with the concentrations of these structures spanning four orders of magnitude. The number of HMO structures found in the milk of a particular mother is often more than 100, although the profile of the structures varies between mothers[53,54]. HMOs contain varying amounts of Lewis-blood-group antigens Le^a, Le^b, Le^x and Le^y (ref. 58). Individuals who produce Le^b epitopes (α(1,2)-fucosylated structures) in their secretions, due to the presence of an active fucosyltransferase 2 (*FUT2*) gene, are known as secretors[59]. Secretors tend to have higher amounts of HMOs than do non-secretors (as much as 20% more). They also produce higher levels of fucosylated structures (nearly twofold more). However, non-secretors often have higher levels of sialylated structures[57] (Fig. 1a). The percentage of non-secretors varies geographically: they comprise about 20% of the population in Europe and up to 40% in West Africa[57].

Whether the HMO profiles of breast milk change as a function of time after delivery, and how differences in HMO composition relate to the development of the gut microbiota and healthy growth of the infant are important unanswered questions. Nanoflow liquid chromatography mass spectrometry has only been applied to HMO profiling during the past several years and the assay imposes constraints on throughput[60]. Limited information is therefore available about how specific HMOs change with time in healthy mothers, and whether consistent differences exist in the HMO profiles across groups of women representing different ages, parities, geographic locations, nutritional states, culinary traditions and socio-economic statuses. The general trend during lactation is a decrease in levels of total HMOs as the mother progresses from the production of colostrum to mature milk, with the largest drop occurring in the first month postpartum[61]. However, the total amount of milk that is delivered as colostrum is quite small compared with the volume of mature milk (matching the size of the infant's stomach and intestine). Therefore, throughout lactation, the amount of each class of HMO — and even each specific HMO structure — provided to the infant remains relatively constant[61–63].

Giving birth prematurely can significantly affect the profile of a mother's HMO structures[64]. HMO profiles cannot yet be predicted in these mothers. Many mothers who deliver preterm have fucosylated HMOs that are as low as 20–40% of total HMOs, but some have levels of greater than 60%. This discrepancy is not corrected over time.

A study published this year demonstrated that the HMO content of Malawian mothers' milk correlates with infant growth outcomes[65]. Breast milk samples collected at 6 months postpartum were divided into two groups: those from mothers whose infants exhibited healthy growth at the time of collection (as defined by anthropometry), and those from mothers whose offspring exhibited severe stunting. Liquid chromatography–time-of-flight mass spectrometry revealed that mothers of infants with stunted growth had significantly lower concentrations of total, sialylated and fucosylated HMOs, with the most growth-discriminatory sialylated HMO being sialyllacto-*N*-tetraose b, and the most discriminatory fucosylated HMOs being 2′-fucosyllactose and lacto-*N*-fucopentaose I.

Sialic acids constitute a group of nine-carbon monosaccharides that are derived from neuraminic acid, and include Neu5Ac. UDP-*N*-acetylglucosamine-2-epimerase, which is the rate-limiting enzyme in the biosynthesis of sialic acid, is produced at low levels in the livers of infants[66]. Breast milk is therefore an important source of these sugars. The availability of sialic acids affects many organs, including the brain, where Neu5Ac is a component of gangliosides and is covalently linked to neural cell-adhesion molecules (NCAMs) that mediate cell–cell interactions involved in synaptogenesis and memory[67,68]. Supplementation of the diet with sialylated glycoproteins and sialyllactose increases the polysialylation of NCAM and sialyated gangliosides, with some reports showing improved memory in animal models[69]. A preclinical model has demonstrated that 6′-sialyllactose also increases muscle mass and contractility[70].

Several HMO structures have been produced chemically and

© 2016 Macmillan Publishers Limited. All rights reserved.

enzymatically[71]. However, producing the wide array of structures encountered in human milk is not yet commercially feasible. There is an approximately 25% overlap between bovine milk oligosaccharide and HMO structures. Sialylated oligosaccharides are present in mature human milk at concentrations that are up to 20-fold greater than in mature bovine milk[72,73]. Therefore, bovine-milk-derived infant formulas, as well as complementary or therapeutic foods that are used to treat children with undernutrition, are deficient in these compounds. However, bovine milk oligosaccharides (BMOs) that have structural similarity to HMOs are present in the by-products of dairy processing, providing an opportunity to purify them at a scale sufficient for preclinical and clinical studies, and potentially for wider distribution should such studies demonstrate sufficient safety and efficacy, and yield an understanding of their mechanism of action.

A study of gnotobiotic animals has provided direct evidence that sialylated milk oligosaccharides are causally related to growth[65]. Young germ-free mice and newborn germ-free piglets were colonized with members of the gut microbiota of a Malawian infant who exhibited stunted growth. Recipient animals were fed a diet representative of foods consumed after weaning by Malawians, with or without supplementation with sialylated BMOs that had been purified from a whey waste stream generated during the manufacture of cheese. The study revealed that sialylated BMOs promote lean body-mass gain, improve metabolic flexibility[74] and affect bone growth. These effects were not ascribable to differences in food consumption. They were also microbiota-dependent: they were not observed in germ-free animals. Moreover, growth promotion was not observed when the animals were provided an isocaloric Malawian diet supplemented with a mixture of fructo-oligosaccharides, a component of some infant formulas.

The milk-oriented microbiota

The initial microbiota of nursing infants is an assemblage of microbes derived from mother's faecal, vaginal and skin microbiota[52]. Within weeks, promicrobial and antimicrobial agents in breast milk help to guide development of a milk-oriented microbiota. A common enrichment involves members of the Actinobacteria, mainly *Bifidobacterium* species, that frequently dominate the gut microbiota of breastfed infants, in some cases representing 70–90% of the faecal community[75]. Intriguingly, this enrichment is less pronounced in infants from more industrialized countries[75–78]. Bifidobacterial enrichment is linked to maternal genotype; the breast milk of secretors seems to enrich bifidobacteria more rapidly[76].

Several beneficial functions have been attributed to a milk-oriented microbiota that is dominated by bifidobacteria. For example, lactate and acetate, the primary end products of bifidobacterial fermentation, are important sources of energy for colonocytes. They also lower intestinal pH and contribute to gut barrier function[79]. Robust colonization by a single bifidobacterial subspecies, *Bifidobacterium longum* subsp. *infantis*, correlates with improved vaccine responses during the first year of life[77]. Intestinal bifidobacteria also produce essential nutrients, including folate and riboflavin[80].

Two dominant species of *Bifidobacterium*, *B. longum* and *B. breve*, routinely colonize breastfed infants throughout the world, although other species, including *B. bifidum*, *B. catenulatum* and *B. pseudocatenulatum* are also commonly observed. In general, bifidobacteria are prolific consumers of HMOs; they possess an array of glycoside hydrolases (notably fucosidases and sialidases[81]) that catalyze the cleavage of key glycosidic linkages, permitting metabolism of some or all of the sugar monomers that are embedded in HMOs. The mechanisms for HMO consumption by these organisms follow two different strategies[82]. *B. longum* subsp. *infantis* and, to a lesser extent, *B. longum* subsp. *longum*, *B. breve* and *B. pseudocatenulatum*, transport HMOs directly into the cell through ATP-binding cassette (ABC) transporters and cleave these oligosaccharides with intracellular glycoside hydrolases (Fig. 1b, left)[83]. By contrast, *B. bifidum* deploys glycoside hydrolases to the cell wall for extracellular cleavage of HMOs before importing selected products of degradation. Similarly, *Bacteroides* species, another important set of HMO consumers (and frequent members of milk-oriented microbiota), also deploy external glycoside hydrolases to degrade these structures before they are internalized (Fig. 1b, right)[83].

The 'internalize, then degrade' approach for HMO consumption adopted by the majority of infant-borne bifidobacteria can be viewed as an ingenious strategy for protecting the neonate. These bacteria prevent growth of competitor strains by simple sequestration of available sugar substrates in the colon, a concept consistent with the inverse correlation observed between faecal HMO concentrations and the level of bifidobacterial colonization[76,84]. An important consideration is whether there are deleterious consequences of harbouring a milk-oriented microbiota that is dominated by bacteria that degrade HMO externally. An antibiotic-treated mouse model has been used to show that mucins, large glycoproteins that contain structures similar to those of HMOs, can be externally degraded by *Bacteroides* spp. to release fucose and sialic acid monomers that cross-feed various pathogenic bacteria[85]. External degradation of HMOs could lead to growth of pathogens or pathobionts in the low-diversity neonatal gut microbiota.Three recent studies point to this potential risk. In gnotobiotic mice that were colonized with the microbiota from a Malawian infant with stunted growth, external degradation of sialylated BMOs by *Bacteroides fragilis* released the constituent monosaccharides, including sialic acid, that cross-fed *Escherichia coli* populations[65]. Others have observed *Bacteroides* cross-feeding Enterobacteriaceae in mice that are fed sialyllactose (an oligosaccharide common to mammalian milks) and in nursing piglets[86,87]. Enterobacteriaceae are considered by some researchers to be a harbinger of dysbiosis[88].

These findings suggest that the potential for bacterial cross-feeding on HMOs may be a risk factor for neonates. They also illustrate the extreme caution that should be afforded when composing diets for neonates that harbour low-diversity gut microbiota during early stages of community development. In cases in which a single oligosaccharide prebiotic is being considered, such as fucosyllactose or sialyllactose in infant formula, it would help to know the composition of the infant milk-oriented microbiota to avoid potential cross-feeding of enteropathogens. Alternatively, this problem might be alleviated by the use of synbiotics (a combination of pre- and probiotics) in which the probiotic component is known to readily consume the oligosaccharides provided or derived monomers.

Several challenging questions need to be addressed. First, we know very little about the functions of various HMO structures or why mammalian evolution has produced such a diverse repertoire. Even more diversity could exist given the number of possible glycosidic linkages, suggesting that observed HMO structures were selected for by evolution. Second, we need to better characterize the interactions and relative effect sizes of the antimicrobial and promicrobial components of breast milk on development of the milk-oriented microbiota. One approach for addressing these questions is to use gnotobiotic animals colonized with milk-oriented microbiota from infants representing different gestational ages, milk-feeding histories and growth phenotypes. Alternatively, gnotobiotic animals could be colonized with defined collections of cultured bacterial strains, recovered from a given donor's microbiota; these clonally arrayed collections can be manipulated so that all members, or subsets of members, are added — with or without pathogens and pathobionts — to recipient animals (Fig. 2). Gnotobiotic recipients colonized with these communities can be fed breast milk or infant formula supplemented with defined milk oligosaccharide structures. (Many of the antimicrobial elements of breast milk, including antibodies, lactoferrin and lysozyme, are absent from such formulas.) These models represent one way for determining the rules that govern early phases of development of the human gut microbiota.

The weaning-oriented microbiota and beyond

Culture-independent studies have characterized a programme of gut microbial community development that is executed during the first

© 2016 Macmillan Publishers Limited. All rights reserved.

2–3 years of postnatal life, as infants move from a diet dominated by milk through a period of complementary feeding to a fully weaned state. In one study, monthly collection of faecal samples from members of a Bangladeshi birth cohort with healthy growth phenotypes allowed the generation of 16S rRNA-sequence-based data sets that described the bacterial composition of their developing gut communities[78]. This study used Random Forests-based models to identify a group of age-discriminatory bacterial strains, the relative abundances of which defined the state of development ('age') of a child's microbiota. Remarkably, many of these age-discriminatory strains were also present in models of normal microbiota development in healthy Malawian infants and children[89].

Deviations from normal can be expressed in the form of a microbiota-for-age *Z*-score (MAZ). Calculating MAZ scores disclosed that microbiota development was impaired in Malawian and Bangladeshi children who presented with moderate or severe acute malnutrition[78,89]. Their microbial communities appeared 'younger' than would be expected from their chronological age. Moreover, this microbiota immaturity is not durably repaired by treatment with current ready-to-use therapeutic foods[78,89]. Transplanting immature microbiota from Malawian children who are stunted or underweight, or from chronologically age-matched donors who have healthy growth phenotypes, to young germ-free mice fed a diet resembling that consumed by the microbiota donors showed that immature microbiota transmit impaired growth phenotypes[89].

These and other studies provide preclinical proof-of-concept that gut microbiota development is causally related to healthy growth[89,90]. They also provide a microbial measure of normal as well as perturbed postnatal development. An important question is how microbiota development affects development of the immune system. This issue can be addressed in part by defining IgA responses of the gut mucosa to members of the microbiota[91], using faecal samples serially collected from members of birth-cohort studies[92]. This approach represents one way to identify relationships between microbial community development, development of the immune system, breast milk HMO content and host growth phenotypes.

A call for human microbial community observatories

Characterizing normal gut microbiota development and the development of other body-habitat-associated microbial communities in members of birth cohorts provides a framework for exploring the degree to which these processes vary across populations of infants and children with healthy growth phenotypes. Whether —and how — perturbations of these programmes are related to growth faltering and the risk for and development of various diseases can also be investigated. These studies should include an examination of the mother and her microbial communities starting at the time of conception, and of the impact of these communities on fetal development. The results could yield insights about as yet unappreciated microbial contributions to a wide range of disorders that are overtly manifest, or foreshadowed, by changes of microbial community structure and function in infancy or childhood (for example, obesity[93–95], immunological disorders including atopic states[96] and neurodevelopmental disorders[97]).

Given the dramatic, myriad and rapid changes in our lifestyles wrought by globalization, as well as the vast differences in sanitation and hygiene experienced by various populations, we propose that a series of 'human microbial observatories' be established to characterize the evolution of microbial communities in mothers before, during and after pregnancy, to monitor fetal development and to characterize the development of microbial communities in their offspring (and perhaps in the future, in the pregnancies of these children and their offspring). We propose that the populations selected for study should not only illustrate currently distinct lifestyles and geographies, but also contain segments that are likely to undergo lifestyle changes within a generation.

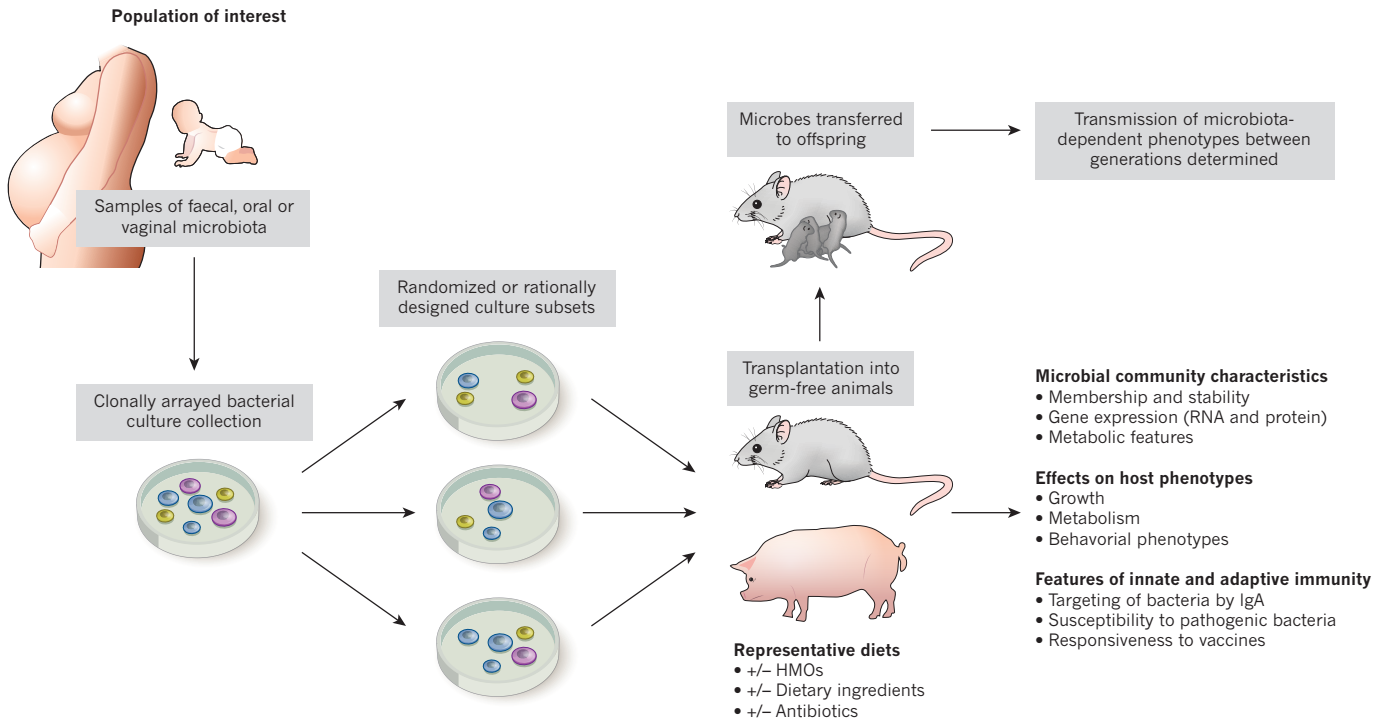


Figure 2 | Discovery pipeline for characterizing the functional properties of developing human microbial communities. Samples of intact, uncultured microbiota are obtained from infants and children with healthy growth phenotypes and normal microbial community development and from those with perturbed community development, or from their mothers. Clonally arrayed collections of cultured organisms are then generated from these microbial communities. The effects of different community configurations on host biology are tested by transplanting these collections, or subsets of the collections, into germ-free animals (mice or other species). Recipient animals are fed diets representative of those consumed by their microbiota donors, or diets designed to test hypotheses about the role of various components, including HMOs, on microbiota-mediated functions. Follow-up studies can be performed by assessing the transmission of microbial communities of interest and associated phenotypes to the offspring of these gnotobiotic animals.

© 2016 Macmillan Publishers Limited. All rights reserved.

Organizations, both private and public, that are committed to addressing global health challenges have already made investments that have enabled durable, trusting relationships to be established between health-care providers and such populations, as well as the infrastructure required to obtain informed consent and apply validated procedures for collecting and archiving biospecimens and associated metadata. Examples include the Global Enteric Multicenter Study (GEMS)[98]; the Etiology, Risk Factors, and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health (MAL-ED) Study[99]; and various water, sanitation and hygiene (WASH) programmes[100]. These investments should be leveraged for the proposed human microbial observatories, which will require sustained support given the extended period of observation. Effective and innovative strategies for achieving such durable support require expertise from multiple disciplines. In our opinion, the development of these strategies is a compelling challenge whose solutions have broad implications for obtaining answers to this biological question as well as myriad others related to the promotion of human flourishing (eudaimonia) in the broadest sense.

Wise and effective stewardship of human microbial resources is a responsibility that extends across generations and national boundaries. Knowledge of how microbial communities evolve in health and how their development is jeopardized or overtly disrupted provides an opportunity to discover strategies and tools for their timely repair. However, understanding how such repair can be achieved brings great responsibility. The immediate as well as long-term consequences of such interventions applied early in the course of a human life need to be determined. Rigorous tests of safety and efficacy have to be designed and applied in representative animal models when available. Thoughtful consideration must be given to the ethical, regulatory and societal issues and consequences that could arise from early interventions that shape the composition and function of our microbial communities. This is a time for inspiration and awe as we gain insight about how we function as holobionts. It is also a time for mindfulness and sobriety as we consider how to deliberately shape facets of our own developmental biology to improve wellness during our human lifecycle. ■

Received 2 November 2015; accepted 25 April 2016.

1. Sender, R., Fuchs, S. & Milo, R. Are we really vastly outnumbered? Revisiting the ratio of bacterial to host cells in humans. *Cell* **164,** 337–340 (2016).
2. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
3. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
4. Gordon, J., Knowlton, N., Relman, D. A., Rohwer, F. & Youle, M. Superorganisms and holobionts. *Microbe* **8,** 152–153 (2013).
5. Levison, M. E., Corman, L. C., Carrington, E. R. & Kaye, D. Quantitative microflora of the vagina. *Am. J. Obstet. Gynecol.* **127,** 80–85 (1977).
6. Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4680–4687 (2011).
7. Dareng, E. O. *et al.* Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. *Epidemiol. Infect.* **144,** 123–137 (2016).
8. Koren, O. *et al.* A guide to enterotypes across the human body: meta-analysis of microbial community structures in human microbiome datasets. *PLoS Comput. Biol.* **9,** e1002863 (2013).
9. Hillier, S. L. *et al.* Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. *N. Engl. J. Med.* **333,** 1737–1742 (1995).
10. Horner-Devine, M. C. & Bohannan, B. J. Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87,** S100–S108 (2006).
11. Stowell, S. R. *et al.* Microbial glycan microarrays define key features of host–microbial interactions. *Nature Chem. Biol.* **10,** 470–476 (2014).
12. Romero, R. *et al.* The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. *Microbiome* **2,** 18 (2014).
13. Romero, R. *et al.* The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* **2,** 4 (2014).
14. DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl Acad. Sci. USA* **112,** 11060–11065 (2015). **This study showed that the composition of the vaginal microbiota early in pregnancy may predict subsequent premature birth, which raises questions about how this community of microbes shapes maternal health and pregnancy outcomes.**
15. Aagaard, K. *et al.* A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *PLoS ONE* **7,** e36466 (2012).
16. Wilson, C. S. Nutritionally beneficial cultural practices. *World Rev. Nutr. Diet.* **45,** 68–96 (1985).
17. Bisanz, J. E. *et al.* Microbiota at multiple body sites during pregnancy in a rural Tanzanian population and effects of moringa-supplemented probiotic yogurt. *Appl. Environ. Microbiol.* **81,** 4965–4975 (2015).
18. MacIntyre, D. A. *et al.* The vaginal microbiome during pregnancy and the postpartum period in a European population. *Sci. Rep.* **5,** 8988 (2015).
19. Huang, Y. E. *et al.* Homogeneity of the vaginal microbiome at the cervix, posterior fornix, and vaginal canal in pregnant Chinese women. *Microb. Ecol.* **69,** 407–414 (2015).
20. Burke, B. S. & Stevenson, S. S. Nutrition studies during pregnancy; relation of maternal nutrition to condition of infant at birth; study of siblings. *J. Nutr.* **38,** 453–467 (1949).
21. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150,** 470–480 (2012).
22. Liu, B. *et al.* Deep sequencing of the oral microbiome reveals signatures of periodontal disease. *PLoS ONE* **7,** e37919 (2012).
23. Duran-Pinedo, A. E. *et al.* Community-wide transcriptome of the oral microbiome in subjects with and without periodontitis. *ISME J.* **8,** 1659–1672 (2014).
24. Siqueira, F. M. *et al.* Intrauterine growth restriction, low birth weight, and preterm birth: adverse pregnancy outcomes and their association with maternal periodontitis. *J. Periodontol.* **78,** 2266–2276 (2007).
25. DiGiulio, D. B. *et al.* Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: a molecular and culture-based investigation. *PLoS ONE* **3,** e3056 (2008).
26. DiGiulio, D. B. *et al.* Prevalence and diversity of microbes in the amniotic fluid, the fetal inflammatory response, and pregnancy outcome in women with preterm pre-labor rupture of membranes. *Am. J. Reprod. Immunol.* **64,** 38–57 (2010).
27. Han, Y. W. *et al.* Transmission of an uncultivated *Bergeyella* strain from the oral cavity to amniotic fluid in a case of preterm birth. *J. Clin. Microbiol.* **44,** 1475–1483 (2006).
28. Han, Y. W., Shen, T., Chung, P., Buhimschi, I. A. & Buhimschi, C. S. Uncultivated bacteria as etiologic agents of intra-amniotic inflammation leading to preterm birth. *J. Clin. Microbiol.* **47,** 38–47 (2009).
29. Swati, P., Thomas, B., Vahab, S. A., Kapaettu, S. & Kushtagi, P. Simultaneous detection of periodontal pathogens in subgingival plaque and placenta of women with hypertension in pregnancy. *Arch. Gynecol. Obstet.* **285,** 613–619 (2012).
30. Costalonga, M. & Herzberg, M. C. The oral microbiome and the immunobiology of periodontal disease and caries. *Immunol. Lett.* **162,** 22–38 (2014).
31. Teng, F. *et al.* Prediction of early childhood caries via spatial–temporal variations of oral microbiota. *Cell Host Microbe* **18,** 296–306 (2015).
32. Kustner, O. Beitrag zur Lehre von der puerperalen Infection der Neugeborenen. *Arch. Gynakol.* **11,** 256–263 (1877).
33. Harris, J. W. & Brown, J. H. The bacterial content of the uterus at cesarean section. *Am. J. Obstet. Gynecol.* **13,** 133–143 (1927).
34. Benirschke, K. Routes and types of infection in the fetus and the newborn. *AMA J. Dis. Child.* **99,** 714–721 (1960).
35. Verstraelen, H. *et al.* Characterisation of the human uterine microbiome in non-pregnant women through deep sequencing of the V1–2 region of the 16S rRNA gene. *PeerJ* **4,** e1602 (2016).
36. Mitchell, C. M. *et al.* Colonization of the upper genital tract by vaginal bacterial species in nonpregnant women. *Am. J. Obstet. Gynecol.* **212,** 611.e1–611.e9 (2015).
37. Stout, M. J. *et al.* Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *Am. J. Obstet. Gynecol.* **208,** 226.e1–226.e7 (2013).
38. Aagaard, K. *et al.* The placenta harbors a unique microbiome. *Sci. Transl. Med.* **6,** 237ra65 (2014).
39. Bhola, K. *et al.* Placental cultures in the era of peripartum antibiotic use. *Aust. N. Z. J. Obstet. Gynaecol.* **48,** 179–184 (2008).
40. Kliman, H. J. Comment on "The placenta harbors a unique microbiome". *Sci. Transl. Med.* **6,** 254le4 (2014).
41. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12,** 87 (2014).
42. Prince, A. L. *et al.* The placental microbiome is altered among subjects with spontaneous preterm birth with and without chorioamnionitis. *Am. J. Obstet. Gynecol.* **214,** 627.e1–627.e16 (2016).
43. Kim, M. J. *et al.* Widespread microbial invasion of the chorioamniotic membranes is a consequence and not a cause of intraamniotic infection. *Lab. Invest.* **89,** 924–936 (2009).
44. Hansen, R. *et al.* First-pass meconium samples from healthy term vaginally-delivered neonates: an analysis of the microbiota. *PLoS ONE* **10,** e0133320 (2015).
45. Ardissone, A. N. *et al.* Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS ONE* **9,** e90784 (2014).
46. Romero, R. *et al.* The role of infection in preterm labour and delivery. *Paediatr. Perinat. Epidemiol.* **15** (suppl. 2), 41–56 (2001).
47. Bearfield, C., Davenport, E. S., Sivapathasundaram, V. & Allaker, R. P. Possible association between amniotic fluid micro-organism infection and microflora in the mouth. *BJOG* **109,** 527–533 (2002).
48. Menon, R., Peltier, M. R., Eckardt, J. & Fortunato, S. J. Diversity in cytokine response to bacteria associated with preterm birth by fetal membranes. *Am. J. Obstet. Gynecol.* **201,** 306.e1–306.e6 (2009).

© 2016 Macmillan Publishers Limited. All rights reserved.

49. Gomez de Agüero, M. *et al.* The maternal microbiota drives early postnatal innate immune development. *Science* **351,** 1296–1302 (2016).
A germ-free mouse model of transient microbial colonization demonstrates that exposure of the mother to microbes during pregnancy shapes immunological development and function in the neonate.
50. Muglia, L. J. & Katz, M. The enigma of spontaneous preterm birth. *N. Engl. J. Med.* **362,** 529–535 (2010).
51. McGuire, M. K. & McGuire, M. A. Human milk: mother nature's prototypical probiotic food? *Adv. Nutr.* **6,** 112–123 (2015).
52. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5,** e177 (2007).
53. Wu, S., Tao, N., German, J. B., Grimm, R. & Lebrilla, C. B. Development of an annotated library of neutral human milk oligosaccharides. *J. Proteome Res.* **9,** 4138–4151 (2010).
54. Wu, S., Grimm, R., German, J. B. & Lebrilla, C. B. Annotation and structural analysis of sialylated human milk oligosaccharides. *J. Proteome Res.* **10,** 856–868 (2011).
55. Kunz, C. & Rudloff, S. Biological functions of oligosaccharides in human milk. *Acta Paediatr.* **82,** 903–912 (1993).
56. Ninonuevo, M. R. *et al.* A strategy for annotating the human milk glycome. *J. Agric. Food Chem.* **54,** 7471–7480 (2006).
57. Totten, S. M. *et al.* Comprehensive profiles of human milk oligosaccharides yield highly sensitive and specific markers for determining secretor status in lactating mothers. *J. Proteome Res.* **11,** 6124–6133 (2012).
58. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22,** 1147–1162 (2012).
59. Thurl, S. *et al.* Variation of human milk oligosaccharides in relation to milk groups and lactational periods. *Br. J. Nutr.* **104,** 1261–1271 (2010).
60. Totten, S. M. *et al.* Rapid-throughput glycomics applied to human milk oligosaccharide profiling for large human studies. *Anal. Bioanal. Chem.* **406,** 7925–7935 (2014).
This paper highlights nanoflow liquid chromatography mass spectrometry, a method that allows the rapid and reproducible detection of HMOs in low-volume biological samples, enabling large-scale clinical studies.
61. Coppa, G. V. *et al.* Changes in carbohydrate composition in human milk over 4 months of lactation. *Pediatrics* **91,** 637–641 (1993).
62. Niñonuevo, M. R. *et al.* Daily variations in oligosaccharides of human milk determined by microfluidic chips and mass spectrometry. *J. Agric. Food Chem.* **56,** 618–626 (2008).
63. Chaturvedi, P. *et al.* Fucosylated human milk oligosaccharides vary between individuals and over the course of lactation. *Glycobiology* **11,** 365–372 (2001).
64. De Leoz, M. L. *et al.* Lacto-*N*-tetraose, fucosylation, and secretor status are highly variable in human milk oligosaccharides from women delivering preterm. *J. Proteome Res.* **11,** 4662–4672 (2012).
65. Charbonneau, M. R. *et al.* Sialylated milk oligosaccharides promote microbiota-dependent growth in models of infant undernutrition. *Cell* **164,** 859–871 (2016).
Gnotobiotic mouse and piglet models were used to show that sialylated milk oligosaccharides play a causal, microbiota-dependent role in lean body-mass gain, bone growth and metabolism.
66. Gal, B. *et al.* Development changes in UDP-*N*-acetylglucosamine 2-epimerase activity of rat and guinea-pig liver. *Comp. Biochem. Physiol. B* **108,** 13–15 (1997).
67. Wang, B. Sialic acid is an essential nutrient for brain development and cognition. *Annu. Rev. Nutr.* **29,** 177–222 (2009).
68. Wang, B. & Brand-Miller, J. The role and potential of sialic acid in human nutrition. *Eur. J. Clin. Nutr.* **57,** 1351–1369 (2003).
69. Wang, B. *et al.* Dietary sialic acid supplementation improves learning and memory in piglets. *Am. J. Clin. Nutr.* **85,** 561–569 (2007).
70. Yonekawa, T. *et al.* Sialyllactose ameliorates myopathic phenotypes in symptomatic GNE myopathy model mice. *Brain* **137,** 2670–2679 (2014).
71. Chen, X. Human milk oligosaccharides (HMOS): structure, function, and enzyme-catalyzed synthesis. *Adv. Carbohydr. Chem. Biochem.* **72,** 113–190 (2015).
72. Aldredge, D. L. *et al.* Annotation and structural elucidation of bovine milk oligosaccharides and determination of novel fucosylated structures. *Glycobiology* **23,** 664–676 (2013).
73. Sundekilde, U. K. *et al.* Natural variability in bovine milk oligosaccharides from Danish Jersey and Holstein-Friesian breeds. *J. Agric. Food Chem.* **60,** 6188–6196 (2012).
74. Muoio, D. M. Metabolic inflexibility: when mitochondrial indecision leads to metabolic gridlock. *Cell* **159,** 1253–1262 (2014).
75. Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z. & Dominguez-Bello, M. G. The infant microbiome development: mom matters. *Trends Mol. Med.* **21,** 109–117 (2015).
76. Lewis, Z. T. *et al.* Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. *Microbiome* **3,** 13 (2015).
77. Huda, M. N. *et al.* Stool microbiota and vaccine responses of infants. *Pediatrics* **134,** e362–e372 (2014).
78. Subramanian, S. *et al.* Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* **510,** 417–421 (2014).
This study used a machine-learning approach to define normal microbiota development in Bangladeshi infants and children and revealed a persistent defect in microbiota development in children that exhibit undernutrition.
79. Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469,** 543–547 (2011).
80. Sugahara, H., Odamaki, T., Hashikura, N., Abe, F. & Xiao, J. Z. Differences in folate production by bifidobacteria of different origins. *Biosci. Microbiota Food Health* **34,** 87–93 (2015).
81. Garrido, D., Dallas, D. C. & Mills, D. A. Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications. *Microbiology* **159,** 649–664 (2013).
82. Garrido, D. *et al.* Comparative transcriptomics reveals key differences in the response to milk oligosaccharides of infant gut-associated bifidobacteria. *Sci. Rep.* **5,** 13517 (2015).
83. Marcobal, A. *et al. Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10,** 507–514 (2011).
84. De Leoz, M. L. *et al.* Human milk glycomics and gut microbial genomics in infant feces show a correlation between human milk oligosaccharides and gut microbiota: a proof-of-concept study. *J. Proteome Res.* **14,** 491–502 (2015).
85. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502,** 96–99 (2013).
86. Frese, S. A. & Mills, D. A. Should infants cry over spilled milk? Fecal glycomics as an indicator of a healthy infant gut microbiome. *J. Pediatr. Gastroenterol. Nutr.* **60,** 695 (2015).
87. Frese, S. A., Parker, K., Calvert, C. C. & Mills, D. A. Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome* **3,** 28 (2015).
88. Shin, N. R., Whon, T. W. & Bae, J. W. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* **33,** 496–503 (2015).
89. Blanton, L. V. *et al.* Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science* **351,** aad3311 (2016).
90. Schwarzer, M. *et al. Lactobacillus plantarum* strain maintains growth of infant mice during chronic undernutrition. *Science* **351,** 854–857 (2016).
91. Kau, A. L. *et al.* Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Sci. Transl. Med.* **7,** 276ra24 (2015).
92. Planer, J. D *et al.* Development of the gut microbiota and mucosal IgA responses in twins and gnotobiotic mice. *Nature* **534,** 263–266 (2016).
93. Cox, L. M. *et al.* Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158,** 705–721 (2014).
94. Dogra, S. *et al.* Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. *mBio* **6,** e02419-14 (2015).
95. Cho, I. *et al.* Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488,** 621–626 (2012).
96. Arrieta, M. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med.* **7,** 307ra152 (2015).
97. Goyal, M. S., Venkatesh, S., Milbrandt, J., Gordon, J. I. & Raichle, M. E. Feeding the brain and nurturing the mind: linking nutrition and the gut microbiota to brain development. *Proc. Natl Acad. Sci. USA* **112,** 14105–14112 (2015).
98. Levine, M. M., Kotloff, K. L., Nataro, J. P. & Muhsen, K. The Global Enteric Multicenter Study (GEMS): impetus, rationale, and genesis. *Clin. Infect. Dis.* **55** (suppl. 4), S215–S224 (2012).
99. MAL-ED Network Investigators. The MAL-ED study: a multinational and multidisciplinary approach to understand the relationship between enteric pathogens, malnutrition, gut physiology, physical growth, cognitive development, and immune responses in infants and children up to 2 years of age in resource-poor environments. *Clin. Infect. Dis.* **59** (suppl. 4), S193–S206 (2014).
This paper describes a large, multi-site birth cohort study that includes an effort to serially sample microbial communities in infants to identify correlations between the composition and the development of the microbiota, postnatal growth phenotypes and other facets of health.
100. Ngure, F. M. *et al.* Water, sanitation, and hygiene (WASH), environmental enteropathy, nutrition, and early child development: making the links. *Ann. NY Acad. Sci.* **1308,** 118–128 (2014).

Acknowledgements Work cited from the authors' laboratories was supported in part by grants from the US National Institutes of Health (DK30292, HD061923, AT007079 and AT008759), the Bill & Melinda Gates Foundation, the March of Dimes Foundation and the Thomas C. and Joan M. Merigan Endowment at Stanford University.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: see go.nature.com/28itcop. Readers are welcome to comment on the online version of this paper at go.nature.com/28itcop. Correspondence should be addressed to J.I.G. (jgordon@wustl.edu).

© 2016 Macmillan Publishers Limited. All rights reserved.

REVIEW

doi:10.1038/nature18846

Diet–microbiota interactions as moderators of human metabolism

Justin L. Sonnenburg[1] & Fredrik Bäckhed[2,3]

It is widely accepted that obesity and associated metabolic diseases, including type 2 diabetes, are intimately linked to diet. However, the gut microbiota has also become a focus for research at the intersection of diet and metabolic health. Mechanisms that link the gut microbiota with obesity are coming to light through a powerful combination of translation-focused animal models and studies in humans. A body of knowledge is accumulating that points to the gut microbiota as a mediator of dietary impact on the host metabolic status. Efforts are focusing on the establishment of causal relationships in people and the prospect of therapeutic interventions such as personalized nutrition.

Worldwide, obesity has more than doubled since 1980 according to the World Health Organization. In 2014, more than 1.9 billion adults were overweight, and over 600 million of those people were obese. Obesity results from a positive energy balance, which occurs when the amount of energy ingested exceeds the amount expended, and it is a strong risk factor for other metabolic complications such as type 2 diabetes. Type 2 diabetes is increasing in prevalence in low-income countries, and in 2014, approximately 422 million adults worldwide had diabetes. The condition is characterized by high blood sugar, resistance to insulin and a relative lack of insulin. Insulin resistance is also associated with an increased flux of free fatty acids that contribute to diabetic dyslipidaemia, which is characterized by a high concentration of triglycerides in blood plasma, a low concentration of high-density lipoprotein (HDL) cholesterol and an increased concentration of small, dense low-density lipoprotein cholesterol particles[1]. Dyslipidaemia is one of the major risk factors for cardiovascular disease in people with diabetes. Accordingly, abnormal metabolism of glucose and lipids is the hallmark of metabolic syndrome, which is defined by central (abdominal) obesity and the presence of two or more of four factors — elevated triglycerides, reduced HDL cholesterol, high blood pressure, and increased fasting blood glucose. As governments and health organizations struggle to find solutions to these largely preventable health issues, a rapidly expanding area of research that is focused on the microbes that live within our digestive tract is offering fresh and interesting insights and potential avenues for intervention.

The human gut is a bioreactor with a microbiota that typically encompasses hundreds or thousands of bacterial taxa, which predominantly belong to two phyla: Firmicutes and Bacteroidetes[2–5]. Tremendous strides have been taken over the past decade towards mapping the composition and basic functional attributes of the gut microbiota of people from industrialized countries[3,5]. This ensemble of organisms has coevolved with the human host and complements the coding potential of our own genome with 500-fold more genes[6]. However, the annotation, and consequently the biological function, of many of these remain poorly defined.

The observation that germ-free mice, which lack a microbiota, have reduced adiposity and improved tolerance to glucose and insulin when compared with conventional (colonized) counterparts[7] jump-started a decade of research that focused on the clarification of underlying mechanisms. Germ-free mice are protected from diet-induced obesity when fed a Western-style diet[8–10], which further supports a link between the gut microbiota and the host metabolism. The altered microbiota that is observed in genetically obese mice[11,12] is sufficient to promote increased adiposity in lean mice that receive a microbiota transplant[12], demonstrating that the microbiota contributes to the regulation of adiposity. The importance and generalizability of these initial findings are strengthened by reports of alterations in the gut microbiota of obese people[4,13–15], which confer the obese or adiposity phenotypes when transferred to mice[14,16].

Here, we review the large body of data that is shaping our understanding of how the gut microbiota can alter the absorption, metabolism and storage of calories. Despite broad agreement that gut microbes modify how the human body responds to components of diet to influence metabolism, the mechanisms that underlie this process are exceptionally complex and the data can be difficult to reconcile. The picture that is emerging suggests that obesity is associated with reduced diversity of the gut microbiota[13,17]. Systemic inflammation and microbial metabolites, such as bile acids and short-chain fatty acids, are also commonly implicated. The ability to easily access and reprogramme the composition and function of the microbiota make it an attractive target for intervention.

Diet as an important modulator of the gut microbiota

Extensive research on the gut microbiota has shown that diet modulates the composition and function of this community of microbes in humans and other mammals[18–25], with the earliest literature[26] published almost 100 years ago. Human intervention studies from the past decade have revealed the extent to which different aspects of the microbiota can be influenced through dietary change; this can be summarized by three main themes.

The first theme is that the microbiota of the human gut responds rapidly to large changes in diet. The existence of these fast, diet-induced dynamics is supported by evidence from people who switch between plant- and meat-based diets, who add more than 30 grams per day of specific dietary fibres to their diet or who follow either a high-fibre–low-fat diet or a low-fibre–high-fat diet for 10 days; in all cases, the composition and function of the microbiota shifted over 1–2 days[18,20,23]. Such marked shifts in response to nutrient availability are perhaps unsurprising given that populations of microbes can double within an hour and the gut extensively purges the community every 24–48 hours. This responsiveness might represent an advantageous feature of enlisting

[1]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Wallenberg Laboratory for Cardiovascular and Metabolic Research, Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 413 45 Gothenburg, Sweden. [3]Section for Metabolic Receptology, Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark.

© 2016 Macmillan Publishers Limited. All rights reserved.

microbes as part of the digestive structure — especially when considering the possible day-to-day variation in food that is available to foragers. It might also be an inescapable consequence of dealing with a complex and competitive microbial community that undergoes rapid turnover.

The second theme is that, despite these rapid dynamics, long-term dietary habits are a dominant force in determining the composition of an individual's gut microbiota. Despite detectable responses of the microbiota within 24 hours of dietary intervention, a 10-day feeding study in 10 people[20] failed to alter the major compositional features and the overall classification of each participant's microbiota. Some, but not all, cross-sectional studies reported that long-term dietary trends are linked to features of microbiota composition[20,25,27,28].

The third theme is that a particular change in diet can have a highly variable effect on different people owing to the individualized nature of their gut microbiota. For example, *Ruminococcus bromii*-related taxa bloomed in response to resistant-starch intervention in most of the 14 obese men in one study; the lack of response in the other individuals might reflect an absence of such taxa in those people[23]. A dietary intervention that includes a boosted intake of fibre and a decreased intake of energy can increase microbiota diversity — as defined by the gene content of the faecal metagenome — for individuals who start with a low microbiota gene content, but not those who start with a high gene content[21]. These individualized responses might fit into categories that enable a precision rather than a personalized approach to understanding responsiveness to diet.

The influence of diet on aspects of microbiota function might also help to explain how a specific metabolic input can alter microbiota composition over time. In a study that focused on the enzymatic activity of trimethylamine lyase, mice that harbour microbiotas with low production of trimethylamine (TMA) could be converted into high producers when their diet was supplemented with the TMA-containing compound L-carnitine for 10 weeks[27]. Similarly, a microbiota-encoded degradation system for porphyran, a polysaccharide that is found in certain species of edible seaweed, is rare in the microbiotas of Western people but prominent in those of populations that regularly consume seaweed[29]. This suggests that certain metabolic inputs can select for pathways as well as the organisms that harbour those pathways. One corollary of this interpretation is that there must be a reservoir of selectable functions — either present at low levels within the gut microbial community or able to invade from an environmental source. It is important to note that numerous other non-dietary mechanisms, such as interstrain killing that is mediated by the type VI secretion system, infection with bacteriophages and priority effects of colonization through which strains are able to exclude one another on the basis of relatedness of particular genetic loci, can underlie microbial community dynamics and might interact with or operate in parallel to dietary-mediated effects[30–33].

Several issues can complicate the unravelling of mechanisms and the interpretion of data in dietary intervention studies in humans. People are notoriously poor at adhering to dietary regimes, and it is difficult to accurately measure the extent of their adherence because the self-assessment of food intake can be clouded by numerous factors. Budget limitations often mean that researchers must choose either tightly controlled studies of small cohorts, for example, in which food is provided, or larger cohort studies that could be confounded by the free will of the participants and by their self-assessment. Because dietary change often involves both the elimination and addition (that is, the substitution) of dietary components, even the most successful intervention studies can raise questions about which diet modification was responsible for the change in the microbiota. A further complication is that many of the dietary changes in such studies also have the potential to directly influence host metabolism in a microbiota-independent way.

As an alternative, animal models enable researchers to tightly control the diet of subjects and to have multiple biological replicates that represent the response of a single microbiota. Experimental models that lack a gut microbiota offer further power for determining whether the effects of diet in the host depend on the microbiota. For example, germ-free rats harvest less energy from a polysaccharide-rich diet[34] and germ-free mice have a reduced adiposity despite an increased intake of food by

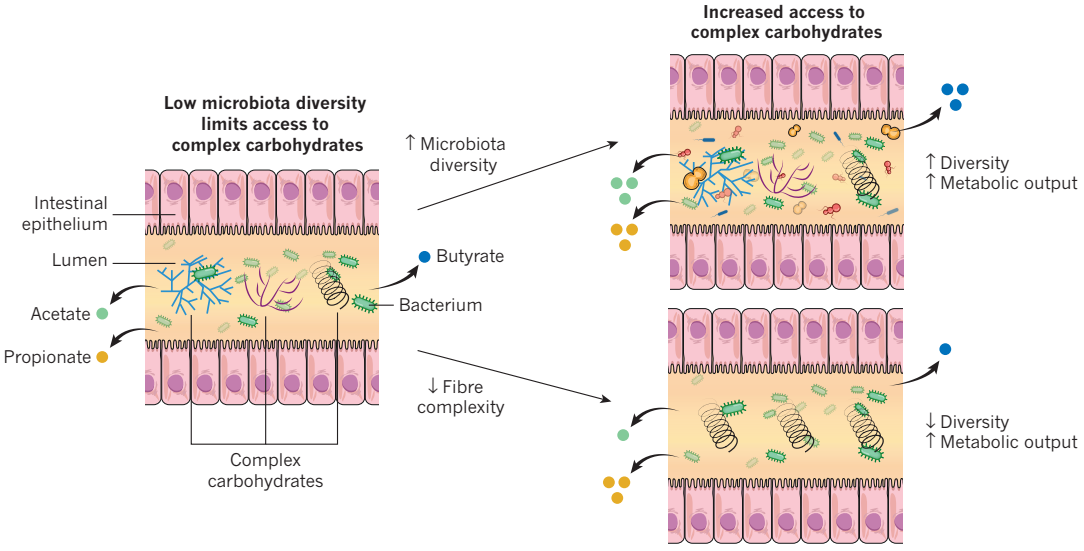


Figure 1 | Interactions between the diet and the gut microbiota dictate the production of short-chain fatty acids. Dietary fibre is a source of complex carbohydrates, which are required for the production of short-chain fatty acids such as acetate, butyrate and propionate. When the diversity of the microbiota is high and the diet contains many types of complex carbohydrates (top right), a relatively high percentage of complex carbohydrates will be accessible to the microbiota. But when the diversity of the microbiota is low and the diet contains many types of complex carbohydrates (left), only a low percentage of these complex carbohydrates are accessible to the microbiota. If the fibre composition of the diet is matched to the needs of a low-diversity microbiota (bottom right) by limiting the types of complex carbohydrate that are available, the levels of production of certain short-chain fatty acids, such as propionate, might increase. However, the diversity of the microbiota will probably remain low and it might not be able to provide as many functions as a diverse microbiota. Consumption of a complex diet (top right) might result in increased levels of production of multiple types of short-chain fatty acids and helps to recruit additional diversity to the gut microbiota. The level of propionate production is correlated with the abundance of *Bacteroides* species in the gut, which is consistent with the involvement of these bacteria in the production of propionate[125]. Fermentation of fibre in the colon has been shown to decrease pH levels, which can help to increase the diversity of the gut microbiota or results in the reinforcement by certain taxa of a pH that favours their own growth[126–129].

© 2016 Macmillan Publishers Limited. All rights reserved.

comparison with their colonized counterparts[7], which demonstrates that the microbiota helps to extract energy from food. These results are consistent with the fact that the fermentation of dietary fibre represents one of the dominant microbial metabolic activities in the colon, the region of the gut in which the microbiota is most dense[35,36].

The short-chain fatty acid end-products of fermentation in the gut can be absorbed into the circulation to serve as both microbiota-generated calories and important regulatory molecules, and it has been estimated that people who consumed a typical British diet in the 1980s received 6–10% of their energy from short-chain fatty acids[37]. By contrast, people who eat large quantities of plants, the main source of dietary fibre[38], such as those in certain African communities that consume up to sevenfold more fibre than people the industrialized world[39], might generate considerably more short-chain fatty acids, which therefore probably contribute more to the whole-body energy requirement. This is in agreement with the increased abundance of taxa that ferment polysaccharides in the gut microbiota of African populations[40]. Certain recurrent physiological states in mammals, such as the non-hibernating period in bears[41] and advanced pregnancy[42], result in a markedly altered microbiota with an increased capacity to harvest energy from the diet without metabolic derangement. It should also be noted that the effects observed in animal models extend beyond a simple improvement in calorie harvest. The microbiota of mice suppress the expression of intestinal angiopoietin-like protein 4, an inhibitor of the enzyme lipoprotein lipase, which increases lipoprotein-lipase activity in adipose tissue and promotes the storage of fat[7]. Accordingly, mice that are deficient in *Angptl4* have increased adiposity, even under germ-free conditions[7].

Experiments that use a Western-style diet, which is devoid of fibres and rich in calories from saturated fat and sucrose, demonstrate that the gut microbiota regulates obesity through additional pathways[8]. For example, germ-free mice are protected from diet-induced obesity when fed high levels of sucrose and lard[8], a diet that alters the composition of the gut microbiota. The presence of the microbiota is both necessary and sufficient for obesity: the transfer of microbiota from mice fed a Western diet to germ-free mice transfers the obese phenotype[43]. By contrast, germ-free mice that are fed a high-fat diet with less sucrose are only partly protected against obesity[44], and all protection from obesity (that is, microbiota-dependent obesity) is lost when sucrose is omitted from the diet[45]. The molecular mechanisms that underpin this finding are unknown. The source of dietary fat also seems to be important. Saturated and unsaturated fats have profoundly different effects on the gut microbiota, and the altered microbiota that results from feeding unsaturated fats can offer protection from lard-induced weight gain[44]. These findings suggest that simple carbohydrates and fats could exert unexpected effects on the host metabolism through the microbiota. Further research is required to clarify how microbial taxa and ecosystems interact with specific macronutrients.

Emerging evidence suggests that the deleterious metabolic effects of processed foods might involve more than just macronutrients. Emulsifiers and artificial sweeteners have been shown to be involved in the development of metabolic syndrome features through their modulation of the microbiota in mice[46,47]. In a study in seven people, artificial sweeteners given at high doses resulted in insulin resistance after only 7 days[47]; however, this dramatic finding needs to be reproduced in a larger study. These data provide evidence that artificial food additives might contribute to metabolic disease through disruption of the microbiota. Notably, an important and unwavering commonality of Western dietary trends is the paucity of plant-based dietary fibre[48], an important fuel for the microbiota. The absence of dietary fibre together with an abundance of nutrients that negatively affect the microbiota could be of considerable importance for understanding metabolic diseases.

Microbial ecology in metabolic disease

The interaction of numerous species, the allocation of resources and the dynamic response to perturbation within the gut provide many of the hallmarks of a complex ecosystem. The application of macroecological concepts to the gut microbiota might therefore be instructive in guiding scientific inquiry and understanding[49], particularly when considering the associations between microbiota diversity and metabolic output (such as the link between short-chain fatty acids and obesity and metabolic disease). For example, many macroecology data suggest that the extent of biodiversity within an ecosystem can serve as an important measure of stability and robustness[50], which are relevant to research that looks at the link between gut microbes and health.

Three metagenomic studies[13,21,51] have shown that improved metabolic health is associated with a relatively high microbiota gene content and with an increased microbial diversity. These data indicate that the extent of the diversity might be an important factor for metabolic health, which is consistent with findings from microbiota studies that have focused on traditional human societies. The gut microbiota of eight hunter–gatherer or rural farming populations in various parts of the world showed increased bacterial diversity compared with those of Western populations[19,40,52–55]. Notably, the microbial taxa that are absent from the Western gut are found in many populations of traditional people that have been separated for thousands of years on different continents. The parsimonious explanation for this is that industrialization has been accompanied by an overall decline in gut microbiota biodiversity as well as the loss of specific phylogenetic groups — a potential consequence of modern lifestyles, medical practices and processed foods. It is unclear whether certain taxa are keystones that promote diversity. It is also unknown whether the increased diversity is only a reflection of a healthy and varied diet or whether it directly contributes to protection from metabolic disease. One theory is that the microbiota of industrialized nations are experiencing a widespread change in functional capacity (for instance, altered production of short-chain fatty acids), which is contributing to modern health issues such as obesity[56–58]. Dietary reinforcement, and specifically the provision of diverse complex carbohydrates, could provide the key to sustaining, and perhaps recovering, a diverse resident ecosystem that is capable of the functions that the human body expects or requires (Fig. 1). A caveat is that diversity can be measured in many ways that include or exclude the relative abundances of species and the functions encoded within them. It is also important to note that a high level of biodiversity does not always correspond to a health-promoting ecosystem: for example, bacterial vaginosis is characterized by a diversity greater than that observed in a healthy state[59]. Undoubtedly, an understanding of diversity within the context of organism identity, location and function enriches the utility of measures that fail to capture important details when used alone.

Fuel for the microbial ecosystem

Many of the plant polysaccharides that are found within dietary fibre are structurally complex. It is therefore unsurprising that the numerous enzymes that are required to de-modify, liberate, transport and metabolize component monosaccharides are not encoded within the human genome[60]. Furthermore, the time that would be required to perform these steps is probably not compatible with the rapid transit that occurs in the small intestine, the region of the gut in which simple carbohydrates are digested and absorbed. Consequently, complex carbohydrates travel to the distal gut for fermentation by its dense community of microbes.

Many complex plant carbohydrates qualify as dietary fibre, according to laboratory tests. However, the amount of fibre that can be metabolized (for example, through the enzymatic degradation of glycosidic linkages and the fermentation of liberated monosaccharides into short-chain fatty acids) will depend on many factors, including the composition of the microbiota. Carbohydrates that can be metabolized by the microbiota are known as microbiota-accessible carbohydrates[61] and can be contrasted with those that pass through the digestive tract without undergoing metabolic transformation. This metabolic accessibility is an important distinguishing characteristic: it defines a carbohydrate as a resource that drives the interspecies economy within the gut and it implies that metabolic products, such as short-chain fatty acids, will be generated.

© 2016 Macmillan Publishers Limited. All rights reserved.

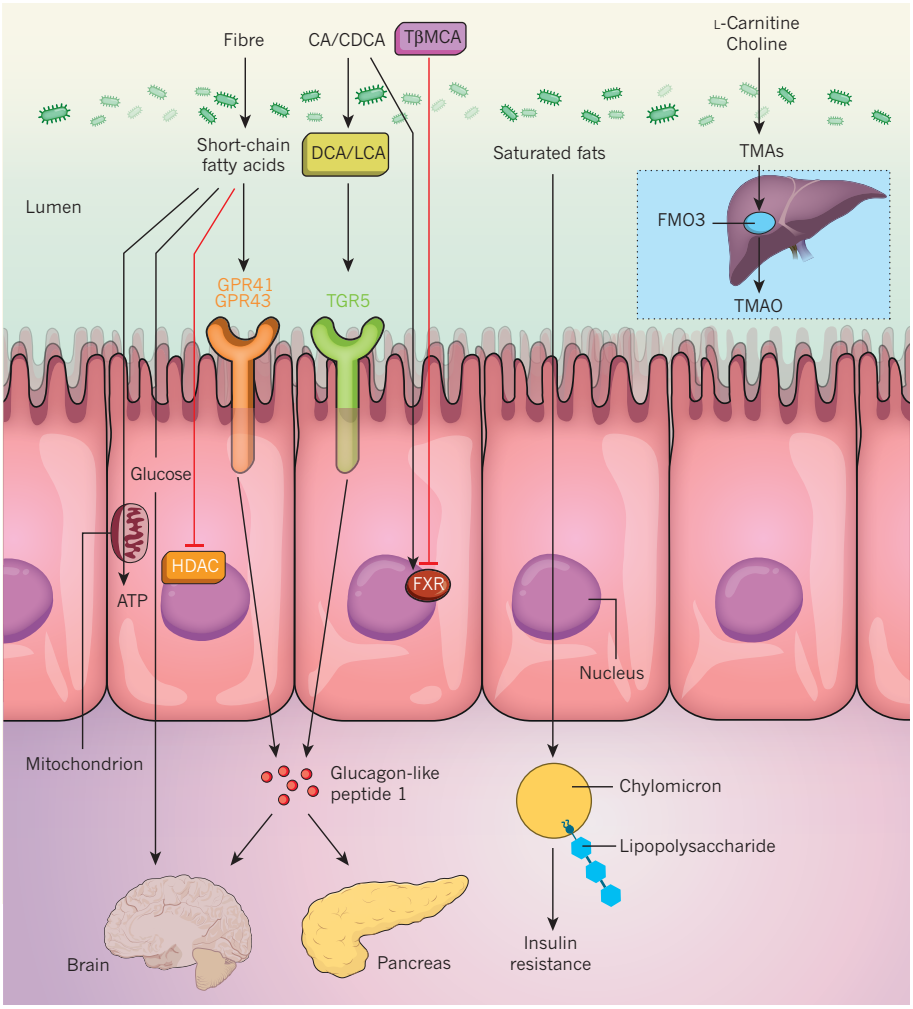


Figure 2 | Mechanisms of signalling from the gut microbiota to the host. The gut microbiota interacts with dietary components and metabolites to form bioactive metabolites that signal to the host through distinct mechanisms. Short-chain fatty acids that are produced by the fermentation of fibre are an important source of energy (ATP) for colonocytes. They are also a substrate for gluconeogenesis, which modulates central metabolism, and are involved in signalling to the host by inhibiting histone deacetylase (HDAC) or by activating G-protein-coupled receptors such as GPR41 and GPR43, which triggers the release of the hormone glucagon-like peptide-1. The primary bile acids cholic acid (CA) and chenodeoxycholic acid (CDCA) are metabolized into the secondary bile acids deoxycholic acid (DCA) and lithocholic acid (LCA), which activates signalling to the host through the G-protein-coupled bile acid receptor 1 (GPBAR1; also known as TGR5). Tauro-β-muricholic acid (TβMCA) is deconjugated into β-muricholic acid (βMCA; not shown), which alleviates the inhibition of the farnesoid X-activated receptor (FXR; also known as the bile acid receptor) by TβMCA. Microbially produced endotoxins (also known as lipopolysaccharides) are taken up into chylomicrons that are formed from dietary saturated fats and subsequently they promote inflammation in the host that induces insulin resistance. L-Carnitine and choline, compounds that are found in red meat, are metabolized into TMAs that are oxidized further into TMAO by the enzyme flavin-containing monooxygenase 3 (FMO3) in the liver (inset).

Notably, high diversity in the microbiota corresponds with high levels of short-chain fatty acid production in rural farmers in Burkina Faso[19], as well as with the enrichment of genes in the microbiome of hunter–gatherers that are associated with the metabolism of complex carbohydrates[62]. In a multigenerational study in mice, the consumption of a Western-style diet exacerbated the loss of microbiota diversity compared with a diet that was rich in microbiota-accessible carbohydrates, and the extinction of taxa corresponded with a predicted loss in diversity of glycoside hydrolases[63]. Several studies in humans indicate that there is a population-specific 'ceiling' on microbiota diversity and metabolic output. For example, following a vegan diet for at least 6 months or a high-fibre–low-fat diet for 10 days were insufficient to substantially increase microbiota diversity or production of faecal short-chain fatty acids[28]. A plant-based diet could significantly alter the composition of the gut microbiota, although a change in diversity was not observed[18]. When fed high levels of resistant starch, individuals who fail to show a bloom in *Ruminococcus bromii* and its relatives also have the highest levels of undigested starch in their stool, which supports the idea that the composition of the microbiota determines whether a carbohydrate is accessible to the microbiota[23]. Overall, these data suggest that the production of short-chain fatty acids is affected by the existing diversity within a microbiota.

Eating whole grains for just 3 days can improve tolerance to glucose in some people, and these 'responders' show an increased representation of specific glycoside hydrolases within the gut microbiome compared with non-responders who received the same dietary intervention[22]. This indicates that the microbiota might need to already have the capacity to degrade certain complex carbohydrates in the diet to reap the potential benefits of microbiota-accessible carbohydrates. Notably, individuals whose microbiota and glucose tolerance respond to a whole-grain intervention tend to consume diets that are higher in fibre. The complex carbohydrates that are associated with whole grains and that were metabolically accessible to the microbiotas of the responders might therefore have been inaccessible to non-responders who also did not routinely consume high-fibre diets.

Microbial metabolites

Microbes that live in the gut continually produce numerous small molecules through primary and secondary metabolic pathways[64], many of which are dependent on the diet of the host. Although some of these compounds are retained within the gut ecosystem, others will be absorbed into the circulation and then chemically modified (that is, co-metabolized) by the host, and eventually secreted in the urine[65]. Much research has focused on short-chain fatty acids, which have been implicated in diverse roles in obesity and metabolic syndrome. Pathways that generate short-chain fatty acids were found to be enriched in metagenomic studies of obesity, and levels of short-chain fatty acids were elevated in overweight or obese people and animal models[12,66,67], which is consistent with these products of microbial fermentation providing extra calories to the host. By contrast, increased levels of the short-chain fatty acid propionate promoted intestinal gluconeogenesis[68] or were associated with the microbiota following gastric bypass[69], which conferred protection from diet-induced obesity on transfer to germ-free recipient mice. The direct delivery of propionate to the colon through propionate-esterified carbohydrate reduced weight gain in a randomized 24-week study of 60 overweight adults[70].

Short-chain fatty acids can signal to the host through at least four distinct pathways (Fig. 2). First, short-chain fatty acids, particularly

© 2016 Macmillan Publishers Limited. All rights reserved.

butyrate, are an energy substrate for colonocytes[71,72], and in response to reduced energy availability, germ-free mice slow down the transit through the small intestine to allow more time for nutrient absorption[73]. Second, propionate is a substrate for gluconeogenesis and can induce intestinal gluconeogenesis, which signals through the central nervous system to protect the host from diet-induced obesity and associated glucose intolerance[68]. Third, butyrate and acetate, another short-chain fatty acid, can act as histone deacetylase inhibitors[74,75]. (Acetate acts in peripheral tissues, in which the concentration of butyrate might not be high enough to exert an effect.) Fourth, short-chain fatty acids signal through G-protein-coupled receptors such as GPR41 (also known as FFAR3) and GPR43 (also known as FFAR2), which affects several important processes that include inflammation[76] and enteroendocrine regulation[77]. However, the generation of short-chain fatty acids is only one aspect of microbial metabolism in the gut.

The microbial metabolism of phosphatidylcholine[78], a phospholipid that is abundant in cheese, seafood eggs and meat, and of L-carnitine[27], an amino acid that is abundant in red meat, produce high levels of TMA. Once it has been absorbed from the gut into the bloodstream, TMA circulates to the liver and is enzymatically oxidized to TMA *N*-oxide (TMAO), a compound that has been associated with poor cardiovascular outcomes in humans and the acceleration of atherosclerosis in mice[27,78,79] (Fig. 2). TMA production serves as an excellent example of the interaction between the diet and the microbiota. For example, microbiotas that are capable of producing TMA make the metabolite only when compounds that contain trimethyl ammonium are present in the diet, and some microbiotas (such as those of vegans) are poor producers of TMA[27], even when precursor compounds are transiently provided through the diet. Together, these data suggest that the microbiota evolves to adapt to specific macronutrients. Many of the experiments that demonstrated the atherogenic nature of TMAO involved supplementing the low-fat diets of animals with the compound. Other metabolites probably contribute to metabolic disease — as supported by evidence from people who have undergone bariatric surgery, a procedure that produces long-term weight loss and improved metabolism and reduces the risk of cardiovascular disease and death[80,81] but that is associated with elevated levels of circulating TMAO[16]. The increased levels of TMAO in such patients might reflect the creation of a more aerobic gut environment that is conducive to generation of this metabolite. It is therefore essential to determine the conditions under which TMAO promotes cardiovascular disease and whether TMAO directly affects cardiovascular disease in humans.

Bile acids, formed by the microbiota from host cholesterol, are another group of metabolites with a profound effect on human health[82]. They are metabolized by the microbiota in the lower part of the small intestine and the colon to generate secondary bile acids[83]. They were originally thought only to act as soaps that solubilize dietary fats to promote their absorption, but over the past two decades, it has become clear that they serve as signalling molecules and bind to distinct receptors such as G-protein-coupled bile acid receptor 1 (also known as TGR5) and the bile acid receptor FXR[84] (Fig. 2). The microbiota regulates TGR5 signalling by producing agonists[85] and FXR signalling by metabolizing antagonists[86]. TGR5 and FXR both have a major impact on host metabolism[84] and, accordingly, an altered microbiota might affect host physiology by modulating the signals that pass through these receptors. The capacity to metabolize tauro-β-muricholic acid, a naturally occurring FXR antagonist[86,87], is essential for the microbiota to induce obesity and steatosis, as well as impaired tolerance to glucose and insulin[87–89]. At least some of this effect is mediated by an altered microbiota[89]. Bariatric surgery is associated with an altered microbiota and metabolism of bile acids[16,90]. Mechanistic links between bile acids and bariatric surgery demonstrate that functional FXR signalling is required for a reduction in body weight and an improvement in glucose tolerance after vertical sleeve gastrectomy[90]. Similarly, TGR5 is required for the improved metabolism of glucose following this procedure. Germ-free mice that received a faecal transplant from people who had undergone Roux-en-Y gastric bypass 10 years earlier gained less fat than did mice that were colonized by microbiota from obese people[16]. Some of the beneficial effects of bariatric surgery might therefore be mediated by the altered microbial metabolism of bile acids, which affects their capacity for signalling. Other mechanisms and metabolites might have equally important roles.

The microbiota produces a vast number of metabolites and much work remains to be done to investigate fully their functions in physiology and pathophysiology. Examples of such metabolites include: ethylphenyl sulfate, which is connected to the exacerbation of autistic behaviour in a mouse model[91]; indole propionic acid, which is linked to improved function of the epithelial barrier in the gut[92]; and indoxyl sulfate and *p*-cresyl sulfate, both of which are associated with poor cardiovascular outcomes in people with uraemia (*p*-cresyl sulfate is also associated with insulin resistance)[93–95]. These metabolites undoubtedly give a glimpse of how this poorly explored universe of molecules can affect the host. The relevance of some of these metabolites in humans is yet to be established. Although several bioactive metabolites are the derivatives of amino acids, neither the effect of the quantity and quality of protein in the diet on metabolite synthesis nor the ensembles of microbial genes that are responsible for metabolite production are well understood.

Inflammation and diet

Obesity and insulin resistance are associated with the increased infiltration of macrophages into and the inflammation of adipose tissue[96,97]. Because the gut microbiota is known to contribute to the obese phenotype, at least in mice, it might also contribute to increased adipose inflammation. A model of adipose inflammation that is dependent on the microbiota but independent of diet is supported by evidence from Swiss Webster mice. While consuming a standard diet, these animals develop a similar amount of adiposity to C57Bl6 common laboratory mice that are fed a high-fat diet for 8 weeks. When germ-free Swiss Webster and C57Bl6 mice are fed their respective adiposity-inducing diets, both exhibit reduced adiposity, lower levels of endotoxins (known as lipopolysaccharides) in the circulation and decreased macrophage infiltration into white adipose tissue, as well as improved metabolism of glucose[44,98]. Obesity in mice is also associated with increased numbers of T cells[99,100] and mast cells[101] and reduced numbers of regulatory T cells[102]. In mouse models, the fermentation of fibre and the generation of short-chain fatty acids seem to promote anti-inflammatory responses both within the gut and systemically through regulatory T cells[103–106]. Although dietary fibre and the production of short-chain fatty acids exert a positive metabolic impact through non-immunological mechanisms in a mouse model of diet-induced obesity[68], it is unclear whether similar interactions that are mediated through the immune compartment contribute to metabolic changes. The supplementation of high-fat diets with fermentable fibres protects mice from obesity and associated diseases[107] but the mechanism that underlies this action remains unclear.

The gut microbiota also interacts with the innate immune system to induce adipose inflammation, and mice that lack Toll-like receptor signalling, through loss of either of the adaptor proteins MyD88 or TRIF (also known as TICAM1), have reduced levels of inflammation in adipose tissue and are protected from insulin resistance that is induced by saturated fatty acids[44]. Mice that are deficient in the gene *Myd88*, but not the gene *Trif*, are protected from diet-induced obesity, which therefore separates obesity from insulin resistance and suggests that they are controlled by different mechanisms. Mice raised in conventional conditions that are fed saturated fatty acids exhibit increased levels of endotoxins in the circulation in comparison to mice that consume polyunsaturated fatty acids. Dietary fat has been demonstrated to increase the amount of endotoxins in the blood plasma of both mice[108] and humans[109], probably by allowing endotoxins to be transported across the epithelium on chylomicrons[110]. These higher levels of endotoxins activate Toll-like receptors in adipose tissue that, in turn, induce the expression of the chemokine CCL2, which is required for macrophage infiltration[44]. The source of dietary fat might

© 2016 Macmillan Publishers Limited. All rights reserved.

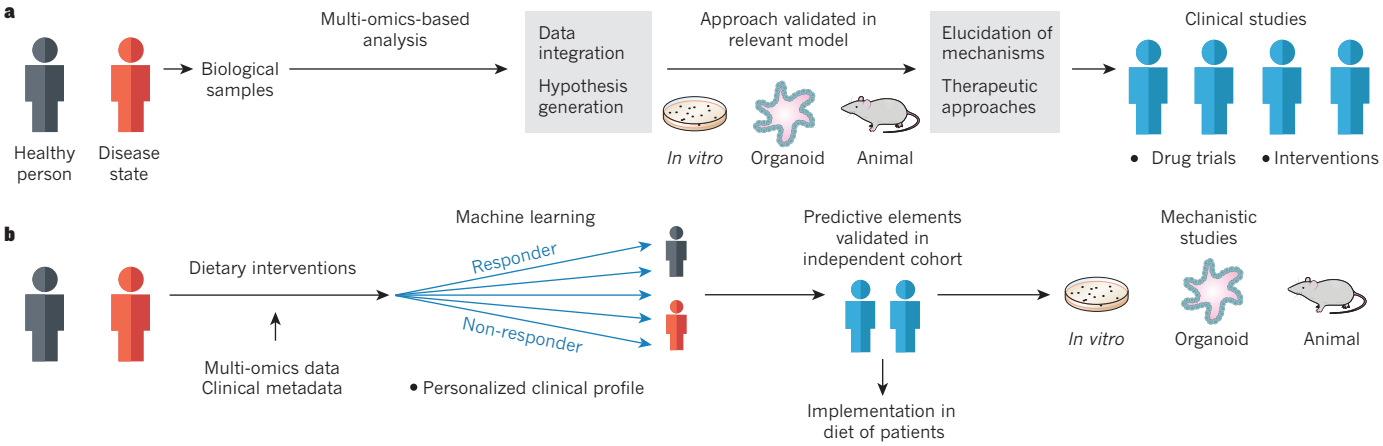


Figure 3 | Strategies for modulating the gut microbiota to improve human health. a, The collection and comparison of multi-omics data from healthy people and those who are affected by metabolic disorders will implicate various genes, pathways and molecules as potential targets for intervention. Relevant experimental models (*in vitro*, organoid or animal models) are then used to elucidate underlying mechanisms and to pilot therapeutic approaches to modulating the gut microbiota, which lay the foundations for intervention studies or drug trials in humans. **b**, Studies in humans can also be a starting point for the identification of strategies to modulate the gut microbiota through components of the diet, which are generally considered to be 'safe' interventions. Data-processing algorithms, such as machine learning, can be used to identify aspects of the clinical profile of individuals (including data on the microbiota) that help to predict the response of others to dietary interventions. After validation of these predictive elements in independent cohorts, the best intervention can be determined and then implemented to improve human health. Such predictive elements can also be used to guide mechanistic studies in experimental models.

therefore have specific interactions with the microbiota that lead to altered interactions with the innate immune system and contribute to metabolic diseases. Mice fed a diet supplemented with fish oil are protected from obesity and insulin resistance. Furthermore, mice that consume lard and receive the microbiota of those fed fish oil are protected against obesity[44], which demonstrates that the modified microbiotas themselves have a protective effect.

A switch to a diet rich in saturated fatty acids shifts the composition of the microbiota[44]. Levels of the bacterium *Bilophila wadsworthia* increase when mice are fed a diet rich in milk fat or supplemented with the bile acid taurocholic acid[111]. Similarly, increased levels of *Bilophila* and a reduced abundance of *Desulfovibrio* were observed in mice that were fed lard compared with fish oil as a source of fat[44]. *B. wadsworthia* increases gut inflammation in mice that lack in the anti-inflammatory cytokine interleukin (IL)-10. Insulin resistance that is induced through a high-fat diet is associated with reduced levels of T helper 17 (T_H17) cells that are positive for IL-17 and retinoic acid receptor-related orphan receptor γt (RORγt)[112]. It is tempting to speculate that one of the underlying mechanisms involves the fat-induced restriction of a specific taxon known as the segmented filamentous bacteria, which induce the expression of IL-23 in enterocytes. IL-23 causes the release of IL-22 from innate lymphoid cells in the ileum, which subsequently induces the production of the proteins serum amyloid A1 and serum amyloid A2 from the epithelium in a paracrine fashion — a process that is required for the activation of T_H17 cells in the ileum[113]. In mice, IL-22 has been shown to protect against metabolic disease, which further suggests a link between the altered gut microbiota, T_H17 cells and IL-22 signalling and the mediation of metabolic disease. However, it is unknown whether taxa that induce specific immune responses, such as the segmented filamentous bacteria in mice, protect against metabolic disease. Despite efforts, there are no reports on the role of segmented filamentous bacteria in people, but other bacteria in the human microbiota might have developed similar functions.

Dietary interventions and diet-based therapeutics

The gut microbiota provides a powerful route to influencing human health. It has many attributes with biomedical potential, such as a connection to multiple facets of human biology, malleability and accessibility for therapeutic targeting or diagnostics. The microbes of the gut can therefore be likened to an easily accessible control centre for the modulation of human physiology. However, owing to the complexity and individuality of each microbiota, the rate at which this potential can be realized is unknown.

Diet and, in particular, polysaccharides serve as primary modulators of the composition and function of the microbiota. Polysaccharides, which are widely consumed components of human food, are therefore functionally analogous to small-molecule drugs. Because of their relative safety (that is, their lack of acute toxicity), availability and low cost, it might be feasible systematically and empirically to determine which dietary polysaccharides, alone or in combinations, can improve human health in different situations.

Such an empirical approach is compatible with emerging concepts in precision health[22,114]. Although the dietary interventions affect the metabolic responses of hosts in an individualized manner, elements of the microbiome can help to predict the response. One study used continuous blood-glucose monitoring to follow postprandial glycaemic responses in 800 people[114]. The responses of individuals to particular foods were highly variable. However, when compared with microbiome profiles and with measurements of metabolism and behaviour, using a machine-learning approach, the response of an individual to a given food could be predicted — even in an independent cohort. Similarly, individuals show large differences in glucose metabolism in response to an intervention that is based on whole grains[22]. Improved tolerance to glucose could be explained largely through enrichment of the genus *Prevotella* within the microbiota. *Prevotella* could also improve the glucose metabolism of mice that were fed carbohydrate-rich diets but not a high-fat diet that was devoid of fermentable polysaccharides. These findings point to the possibility of a mechanism-free, empirical approach for determining a dietary intervention that is appropriate for a given individual or group. They also highlight the potential of a next generation of probiotics (sets of microbiota-derived living microbes that will be tailored to interact with a given diet) as a method for converting non-responders into responders. A further outcome of this approach might be the use of predictive elements of metadata to guide the generation of hypotheses and to determine priorities for investigation into underlying mechanisms.

Perspective

It is becoming clear that an altered gut microbiota is associated with metabolic diseases in humans that range from obesity to type 2 diabetes

© 2016 Macmillan Publishers Limited. All rights reserved.

and cardiovascular disease. Causality has also been demonstrated in animal models. To move forwards, it will be essential to understand whether the gut microbiota is causally linked to host metabolism in humans. Prospective studies should be performed to determine whether the gut microbiota is altered before or after the onset of disease. This will require large cohorts that allow considerable numbers of participants to develop the disease under investigation, and it will probably involve the high-resolution monitoring of host and microbial parameters to determine the progression of derangements.

Another approach is to transfer microbiotas from humans to mice, and this is particularly powerful when focused on twin cohorts to control for human genetics[14,15,51]. In one-such study, transplantation of the microbiota from obese individuals to germ-free mice transfers the obese phenotype, as determined by increased weight gain, whereas administration of *Christensenella minuta* prevents weight gain[15]. In a separate study, bacterial representatives from the microbiota of lean individuals were associated with an increased production of short-chain fatty acids, whereas the microbiota of obese individuals had an increased abundance of genes that are involved in biosynthesis of branched-chain amino acids, which are associated with impaired sensitivity to insulin[14]. Importantly, the lean microbiota could only invade and prevent increased adiposity when the recipient mice consumed a diet that was low in fat and high in fruits and vegetables. Consistent with the idea that the microbiota reinforces the diet, supplementation with *Prevotella* produces an improved tolerance to glucose only when mice are fed a standard diet that is rich in fibre, and not a Western-style diet, which is devoid of fibre[22].

A similar dependency on diet was observed in children with a type of malnutrition known as kwashiorkor[115]. Twins that are discordant for kwashiorkor have distinct microbiotas, and germ-free mice that have been colonized with a 'kwashiorkor' microbiota experience weight loss when they are fed a typical Malawian diet, which is based on tomatoes and corn. However, when the mice are fed a peanut-based, 'ready-to-use' therapeutic food, their weight transiently increases and their microbiota normalize[115]. It is becoming increasingly important to consider how the diet can modify microbiota-linked disease states in mice to generate hypotheses about underlying molecular mechanisms that can then be tested and validated in people. Faecal microbiota transplantation, which has been shown to cure recurrent infection with *Clostridium difficile*[116], has also been used to directly address whether the gut microbiota can affect the metabolism of the host. Eighteen insulin-resistant obese men were randomly designated to receive either an autologous (control) faecal microbiota transplant or a similar transplant from a lean, insulin-sensitive donor. Insulin clamps that were performed before and after the intervention revealed that the insulin sensitivity of a subset of the participants had significantly improved 6 weeks after the transplant[117]. It is unclear whether the positive responses of these individuals are dependent on characteristics of the donors or the recipients as well as what the duration of the responses should be. Research in larger cohorts is required to verify the effects of faecal microbiota transplantion and to answer remaining questions. For example, experiments could be performed using specific bacteria from lean microbiotas with the aim of developing next-generation probiotics. It is clear that stratification might be required to identify groups that are likely to respond to such interventions[22,114].

To improve the understanding of how the microbiota affects the metabolism in humans, metagenomics, transcriptomics, proteomics and metabolomics data from key target tissues and the microbiota during various disease states and interventions should be combined to provide a map of co-occurrences. These data enable the formation of testable hypotheses that can be pursued in validated animal models, and they will form the foundation for precision interventions (Fig. 3).

It will also be important to gain a more nuanced understanding of the foundational principles of the microbiota, such as the cross-sectional or longitudinal spatial organization of interactions between the host and its microbes in the intestine[118]. The majority of studies in humans and mice rely on faecal samples, which provide some representation of what is occurring throughout the digestive tract; however, aspects of microbial communities and host responses that are specific to the small intestine might be obscured by faecal sampling[119,120]. For example, it could miss information on how the microbiota affects nutrient absorption in the small intestine through its impact on glucose transporters and bile acids, which are essential for the absorption of lipids and fat-soluble vitamins.

Microbial metabolites probably act as mediators for the host metabolism and can be either beneficial (for example, butyrate) or detrimental (TMAO). Such molecules might therefore provide fresh therapeutic approaches in which beneficial metabolites could be supplemented pharmacologically or the bacteria that produce them are developed into probiotics. And receptor antagonists could be developed from detrimental metabolites if the relevant receptor has been identified. Another possibility is to target the microbial enzymes that produce metabolites with inhibitors. An inhibitor of TMA lyase that stops the microbial synthesis of TMA and therefore reduces the levels of circulating TMAO prevents the development of atherosclerosis in mice[121]. However, such inhibitors are yet to be tested in humans, and it is unlikely that one metabolite acts alone to promote or prevent metabolic diseases. Strategies that promote or prevent suites of metabolites are more likely to have wider applicability and larger effects on host metabolism.

It is reasonable to consider what proportion of metabolic problems in humans could be addressed by properly caring for the gut microbiota. The use of antibiotics in early life is associated with obesity in both people and mice, which suggests that the disruption of microbial ecosystems at crucial points in time might affect physiology in later life and also that the amendment of medical practices could have a substantial impact[66,122]. However, changes in the diet might be more important for reaping the health benefits that the microbiota can provide. Increased levels of polysaccharides are likely to be of benefit to people who follow a typical Western-style diet, most of whom consume far below the recommended amounts of dietary fibre[48]; meta-analyses show that the increased consumption of fibre significantly decreases the risk of mortality[123,124]. Controlled dietary interventions that document the utility of various supplements, probiotics, nutrients and foods in modulating aspects of the gut microbiota and human health are required. The measurement of multiple aspects of individuality, including the microbiota, will provide insight into the characteristics of people who respond beneficially to a given intervention and will pave the way for microbiota-focused precision nutrition. A deeper understanding of the gut microbiota, an important aspect of failing health, has the potential to contribute big gains in our understanding of metabolic health and weight loss. ■

Received 27 January; accepted 22 April 2016.

1. Mooradian, A. D. Dyslipidemia in type 2 diabetes mellitus. *Nature Clin. Pract. Endocrinol. Metab.* **5,** 150–159 (2009).
2. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308,** 1635–1638 (2005).
3. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
4. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
5. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
6. Li, J. *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nature Biotechnol.* **32,** 834–841 (2014).
7. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101,** 15718–15723 (2004).
8. Bäckhed, F., Manchester, J. K., Semenkovich, C. F. & Gordon, J. I. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc. Natl Acad. Sci. USA* **104,** 979–984 (2007).
9. Rabot, S. *et al.* Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. *FASEB J.* **24,** 4948–4959 (2010).
10. Ding, S. *et al.* High-fat diet: bacteria interactions promote intestinal inflammation which precedes and correlates with obesity and insulin resistance in mouse. *PLoS ONE* **5,** e12191 (2010).
11. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102,** 11070–11075 (2005).
12. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased

© 2016 Macmillan Publishers Limited. All rights reserved.

capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
The first study to show that the microbiota from an obese mouse could confer increased weight gain to a germ-free recipient mouse.
13. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500,** 541–546 (2013).
14. Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341,** 1241214 (2013).
This study showed that a microbiota from a lean individual could invade the microbiota of an obese individual and provide protection from weight gain, but that the invasion and protection was dependent on diet.
15. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159,** 789–799 (2014).
16. Tremaroli, V. *et al.* Roux-en-Y gastric bypass and vertical banded gastroplasty induce long-term changes on the human gut microbiome contributing to fat mass regulation. *Cell Metab.* **22,** 228–238 (2015).
17. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
18. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–563 (2014).
19. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107,** 14691–14696 (2010).
The first of several studies to show that the gut microbiota of a traditional rural population is more diverse than and contains distinct taxa in comparison to the microbiotas of Western populations.
20. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–108 (2011).
21. Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* **500,** 585–588 (2013).
22. Kovatcheva-Datchary, P. *et al.* Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of *Prevotella*. *Cell Metab.* **22,** 971–982 (2015).
23. Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* **5,** 220–230 (2011).
24. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320,** 1647–1651 (2008).
25. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332,** 970–974 (2011).
26. Torrey, J. C. The regulation of the intestinal flora of dogs through diet. *J. Med. Res.* **39,** 415–447 (1919).
27. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Med.* **19,** 576–585 (2013).
28. Wu, G. D. *et al.* Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. *Gut* **65,** 63–72 (2016).
29. Hehemann, J. H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464,** 908–912 (2010).
30. Wexler, A. G. *et al.* Human symbionts inject and neutralize antibacterial toxins to persist in the gut. *Proc. Natl Acad. Sci. USA 113,* 3639–3644 (2016).
31. Chatzidaki-Livanis, M., Geva-Zatorsky, N. & Comstock, L. E. *Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. *Proc. Natl Acad. Sci. USA* **113,** 3627–3632 (2016).
32. Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501,** 426–429 (2013).
33. Reyes, A., Wu, M., McNulty, N. P., Rohwer, F. L. & Gordon, J. I. Gnotobiotic mouse model of phage–bacterial host dynamics in the human gut. *Proc. Natl Acad. Sci. USA* **110,** 20236–20241 (2013).
34. Wostmann, B. S., Larkin, C., Moriarty, A. & Bruckner-Kardoss, E. Dietary intake, energy metabolism, and excretory losses of adult male germfree Wistar rats. *Lab. Anim. Sci.* **33,** 46–50 (1983).
35. Lozupone, C. A. *et al.* The convergence of carbohydrate active gene repertoires in human gut microbes. *Proc. Natl Acad. Sci. USA* **105,** 15076–15081 (2008).
36. El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nature Rev. Microbiol.* **11,** 497–504 (2013).
37. McNeil, N. I. The contribution of the large intestine to energy supplies in man. *Am. J. Clin. Nutr.* **39,** 338–342 (1984).
38. Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70,** 567–590 (1990).
39. Bingham, S. & Cummings, J. H. in *Medical Aspects of Dietary Fiber* (eds Spiller, G. A. & Kay, R. M.) 261–2884 (Plenum, 1980).
40. Schnorr, S. L. *et al.* Gut microbiome of the Hadza hunter–gatherers. *Nature Commun.* **5,** 3654 (2014).
41. Sommer, F. *et al.* The gut microbiota modulates energy metabolism in the hibernating brown bear *Ursus arctos*. *Cell Rep.* **14,** 1655–1661 (2016).
42. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150,** 470–480 (2012).
43. Turnbaugh, P. J., Bäckhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3,** 213–223 (2008).
44. Caesar, R., Tremaroli, V., Kovatcheva-Datchary, P., Cani, P. D. & Backhed, F. Crosstalk between gut microbiota and dietary lipids aggravates WAT inflammation through TLR signaling. *Cell Metab.* **22,** 658–668 (2015).
45. Fleissner, C. K. *et al.* Absence of intestinal microbiota does not protect mice from diet-induced obesity. *Br. J. Nutr.* **104,** 919–929 (2010).
46. Chassaing, B. *et al.* Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. *Nature* **519**, 92–96 (2015).
47. Suez, J. *et al.* Artificial sweeteners induce glucose intolerance by altering the gut microbiota. *Nature* **514,** 181–186 (2014).
48. McGill, C. R., Fulgoni, V. L. III & Devareddy, L. Ten-year trends in fiber and whole grain intakes and food sources for the United States population: National Health and Nutrition Examination Survey 2001–2010. *Nutrients* **7,** 1119–1130 (2015).
49. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. & Relman, D. A. The application of ecological theory toward an understanding of the human microbiome. *Science* **336,** 1255–1262 (2012).
50. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486,** 59–67 (2012).
51. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
52. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).
53. Obregon-Tito, A. J. *et al.* Subsistence strategies in traditional societies distinguish gut microbiomes. *Nature Commun.* **6,** 6505 (2015).
54. Martínez, I. *et al.* The gut microbiota of rural Papua New Guineans: composition, diversity patterns, and ecological processes. *Cell Rep.* **11,** 527–538 (2015).
55. Clemente, J. C. *et al.* The microbiome of uncontacted Amerindians. *Sci. Adv.* **1,** e1500183 (2015).
56. Forslund, K. *et al.* Country-specific antibiotic use practices impact the human gut resistome. *Genome Res.* **23,** 1163–1169 (2013).
57. Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498,** 99–103 (2013).
58. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490,** 55–60 (2012).
59. Srinivasan, S. *et al.* Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLoS ONE* **7,** e37818 (2012).
60. Martens, E. C., Kelly, A. G., Tauzin, A. S. & Brumer, H. The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *J. Mol. Biol.* **426,** 3851–3865 (2014).
61. Sonnenburg, E. D. & Sonnenburg, J. L. Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* **20,** 779–786 (2014).
62. Rampelli, S. *et al.* Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr Biol.* **25,** 1682–1693 (2015).
63. Sonnenburg, E. D. *et al.* Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529,** 212–215 (2016).
64. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science* **349,** 1254766 (2015).
65. Meyer, T. W. & Hostetter, T. H. Uremic solutes from colon microbes. *Kidney Int.* **81,** 949–954 (2012).
66. Cho, I. *et al.* Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488,** 621–626 (2012).
This study demonstrated that the use of antibiotics in early life might cause metabolic disease in later life.
67. Schwiertz, A. *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18,** 190–195 (2010).
68. De Vadder, F. *et al.* Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell* **156,** 84–96 (2014).
69. Liou, A. P. *et al.* Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci Transl. Med.* **5,** 178ra41 (2013).
70. Chambers, E. S. *et al.* Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* **64,** 1744–1754 (2015).
71. Donohoe, D. R. *et al.* The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab.* **13,** 517–526 (2011).
72. Donohoe, D. R., Wali, A., Brylawski, B. P. & Bultman, S. J. Microbial regulation of glucose metabolism and cell-cycle progression in mammalian colonocytes. *PLoS ONE* **7,** e46589 (2012).
73. Wichmann, A. *et al.* Microbial modulation of energy availability in the colon regulates intestinal transit. *Cell Host Microbe* **14,** 582–590 (2013).
74. Thorburn, A. N. *et al.* Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nature Commun.* **6,** 7320 (2015).
75. Davie, J. R. Inhibition of histone deacetylase activity by butyrate. *J. Nutr.* **133,** 2485S–2493S (2003).
76. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461,** 1282–1286 (2009).
77. Samuel, B. S. *et al.* Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc. Natl Acad. Sci. USA* **105,** 16767–16772 (2008).
78. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472,** 57–63 (2011).
79. Tang, W. H. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368,** 1575–1584 (2013).
80. Sjöström, L. *et al.* Effects of bariatric surgery on mortality in Swedish obese subjects. *N. Engl. J. Med.* **357,** 741–752 (2007).
81. Sjöström, L. *et al.* Bariatric surgery and long-term cardiovascular events. *J. Am. Med. Assoc.* **307,** 56–65 (2012).
82. Russell, D. W. The enzymes, regulation, and genetics of bile acid synthesis. *Annu. Rev. Biochem.* **72,** 137–174 (2003).

© 2016 Macmillan Publishers Limited. All rights reserved.

83. Midtvedt, T. Microbial bile acid transformation. *Am. J. Clin. Nutr.* **27,** 1341–1347 (1974).
84. Thomas, C., Pellicciari, R., Pruzanski, M., Auwerx, J. & Schoonjans, K. Targeting bile-acid signalling for metabolic diseases. *Nature Rev. Drug Discov.* **7,** 678–693 (2008).
85. Kawamata, Y. *et al.* A G protein-coupled receptor responsive to bile acids. *J. Biol. Chem.* **278,** 9435–9440 (2003).
86. Sayin, S. I. *et al.* Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR Antagonist. *Cell Metab.* **17,** 225–235 (2013).
87. Li, F. *et al.* Microbiome remodelling leads to inhibition of intestinal farnesoid X receptor signalling and decreased obesity. *Nature Commun.* **4,** 2384 (2013).
88. Jiang, C. *et al.* Intestinal farnesoid X receptor signaling promotes nonalcoholic fatty liver disease. *J. Clin. Invest.* **125,** 386–402 (2015).
89. Parséus, A. *et al.* Microbiota-induced obesity requires farnesoid X receptor. *Gut* http://dx.doi.org/10.1136/gutjnl-2015-310283 (2016).
90. Ryan, K. K. *et al.* FXR is a molecular target for the effects of vertical sleeve gastrectomy. *Nature* **509,** 183–188 (2014).
91. Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155,** 1451–1463 (2013).
92. Venkatesh, M. *et al.* Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. *Immunity* **41,** 296–310 (2014).
93. Meijers, B. K. *et al. p*-Cresol and cardiovascular risk in mild-to-moderate kidney disease. *Clin. J. Am. Soc. Nephrol.* **5,** 1182–1189 (2010).
94. Koppe, L. *et al. p*-Cresyl sulfate promotes insulin resistance associated with CKD. *J. Am. Soc. Nephrol.* **24,** 88–99 (2013).
95. Barreto, F. C. *et al.* Serum indoxyl sulfate is associated with vascular disease and mortality in chronic kidney disease patients. *Clin. J. Am. Soc. Nephrol.* **4,** 1551–1558 (2009).
96. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112,** 1796–1808 (2003).
97. Xu, H. *et al.* Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J. Clin. Invest.* **112,** 1821–1830 (2003).
98. Caesar, R. *et al.* Gut-derived lipopolysaccharide augments adipose macrophage accumulation but is not essential for impaired glucose or insulin tolerance in mice. *Gut* **61,** 1701–1707 (2012).
99. Winer, S. *et al.* Normalization of obesity-associated insulin resistance through immunotherapy. *Nature Med.* **15,** 921–929 (2009).
100. Nishimura, S. *et al.* $CD8^+$ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nature Med.* **15,** 914–920 (2009).
101. Liu, J. *et al.* Genetic deficiency and pharmacological stabilization of mast cells reduce diet-induced obesity and diabetes in mice. *Nature Med.* **15,** 940–945 (2009).
102. Feuerer, M. *et al.* Lean, but not obese, fat is enriched for a unique population of regulatory T cells that affect metabolic parameters. *Nature Med.* **15,** 930–939 (2009).
103. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic T_{reg} cell homeostasis. *Science* **341,** 569–573 (2013).
104. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504,** 446–450 (2013).
105. Trompette, A. *et al.* Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature Med.* **20,** 159–166 (2014).
106. Arpaia, N. *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504,** 451–455 (2013).
107. Cani, P. D. *et al.* Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia. *Diabetologia* **50,** 2374–2383 (2007).
108. Cani, P. D. *et al.* Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56,** 1761–1772 (2007).
The first study to demonstrate that the presence of endotoxin is sufficient to alter glucose metabolism in mice.
109. Erridge, C., Attina, T., Spickett, C. M. & Webb, D. J. A high-fat meal induces low-grade endotoxemia: evidence of a novel mechanism of postprandial inflammation. *Am. J. Clin. Nutr.* **86,** 1286–1292 (2007).
110. Ghoshal, S., Witta, J., Zhong, J., de Villiers, W. & Eckhardt, E. Chylomicrons promote intestinal absorption of lipopolysaccharides. *J. Lipid Res.* **50,** 90–97 (2009).
111. Devkota, S. *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10*$^{-/-}$ mice. *Nature* **487,** 104–108 (2012).
112. Garidou, L. *et al.* The gut microbiota regulates intestinal CD4 T cells expressing RORγt and controls metabolic disease. *Cell Metab.* **22,** 100–112 (2015).
113. Sano, T. *et al.* An IL-23R/IL-22 circuit regulates epithelial serum amyloid A to promote local effector Th17 responses. *Cell* **163,** 381–393 (2015).
114. Zeevi, D. *et al.* Personalized nutrition by prediction of glycemic responses. *Cell* **163,** 1079–1094 (2015).
This study used a machine-learning approach to mine personal health profiles that included microbiome data to predict the postprandial glycaemic response.
115. Smith, M. I. *et al.* Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* **339,** 548–554 (2013).
116. van Nood, E. *et al.* Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* **368**, 407–415 (2013).
117. Vrieze, A. *et al.* Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143,** 913–916 (2012).
This study demonstrated that sensitivity to insulin could be changed by directly altering the gut microbiota through faecal microbiota transplantation.
118. Earle, K. A. *et al.* Quantitative imaging of gut microbiota spatial organization. *Cell Host Microbe* **18,** 478–488 (2015).
119. Lichtman, J. S. *et al.* The effect of microbial colonization on the host proteome varies by gastrointestinal location. *ISME J.* **10,** 1170–1181 (2016).
120. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1,** 6ra14 (2009).
121. Wang, Z. *et al.* Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* **163,** 1585–1595 (2015).
The first example of inhibiting microbial enzymes (or 'drugging the bug') to prevent atherosclerosis.
122. Ajslev, T. A., Andersen, C. S., Gamborg, M., Sorensen, T. I. & Jess, T. Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. *Int. J. Obes.* **35,** 522–529 (2011).
123. Kim, Y. & Je, Y. Dietary fiber intake and total mortality: a meta-analysis of prospective cohort studies. *Am. J. Epidemiol.* **180,** 565–573 (2014).
124. Yang, Y., Zhao, L. G., Wu, Q. J., Ma, X. & Xiang, Y. B. Association between dietary fiber and lower risk of all-cause mortality: a meta-analysis of cohort studies. *Am. J. Epidemiol.* **181**, 83–91 (2015).
125. Salonen, A. *et al.* Impact of diet and individual variation on intestinal microbiota composition and fermentation products in obese men. *ISME J.* **8,** 2218–2230 (2014).
126. Walker, A. W., Duncan, S. H., McWilliam Leitch, E. C., Child, M. W. & Flint, H. J. pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Appl. Environ. Microbiol.* **71,** 3692–3700 (2005).
127. Chung, W. S. *et al.* Modulation of the human gut microbiota by dietary fibres occurs at the species level. *BMC Biol.* **14,** 3 (2016).
128. Bown, R. L., Gibson, J. A., Sladen, G. E., Hicks, B. & Dawson, A. M. Effects of lactulose and other laxatives on ileal and colonic pH as measured by a radiotelemetry device. *Gut* **15,** 999–1004 (1974).
129. Kettle, H., Louis, P., Holtrop, G., Duncan, S. H. & Flint, H. J. Modelling the emergent dynamics and major metabolites of the human colonic microbiota. *Environ. Microbiol.* **17,** 1615–1630 (2015).

Acknowledgements The authors thank members of the Sonnenburg and Bäckhed laboratories for discussions. This work was funded by a grant from the US National Institute of Diabetes and Digestive and Kidney Diseases NIDDK (R01-DK085025 to J.L.S.) and grants from the Swedish Research Council and the Novo Nordisk Foundation to F.B. F.B. is a recipient of a European Research Council Consolidator Grant (615362-METABASE).

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/28j4ikq. Correspondence should be addressed to J.L.S. (jsonnenburg@stanford.edu) and F.B. (fredrik.backhed@wlab.gu.se).

© 2016 Macmillan Publishers Limited. All rights reserved.

REVIEW

doi:10.1038/nature18847

The microbiome and innate immunity

Christoph A. Thaiss[1]*, Niv Zmora[1,2,3]*, Maayan Levy[1]* & Eran Elinav[1]

The intestinal microbiome is a signalling hub that integrates environmental inputs, such as diet, with genetic and immune signals to affect the host's metabolism, immunity and response to infection. The haematopoietic and non-haematopoietic cells of the innate immune system are located strategically at the host–microbiome interface. These cells have the ability to sense microorganisms or their metabolic products and to translate the signals into host physiological responses and the regulation of microbial ecology. Aberrations in the communication between the innate immune system and the gut microbiota might contribute to complex diseases.

The past two decades witnessed a revolution in our understanding of host–microbial interactions that led to the concept of the mammalian holobiont — the result of co-evolution of the eukaryotic and prokaryotic parts of an organism. The revolution required two paradigm shifts that had a tremendous impact on their respective fields. The first occurred during the late 1990s with the discovery of pattern-recognition receptors (PRRs) in the innate immune system that sense microorganisms through conserved molecular structures. Several families of PRRs and their signalling pathways are now known, including the Toll-like receptors (TLRs), the nucleotide-binding oligomerization (NOD)-like receptors (NLRs), the RIG-I-like receptors, the C-type lectin receptors, the absent in melanoma 2 (AIM2)-like receptors and the OAS-like receptors[1]. These sensors are expressed by a variety of cellular compartments and constitute a continuous surveillance system for the presence of microorganisms in tissues.

The second shift occurred fewer than 10 years later and was driven by the culture-independent characterization of the microbiome[2] — the entirety of the microorganisms that colonize the human body and their genomes. Because of the enormous number of microorganisms that reside on the surface of the body — the skin and the gastrointestinal, respiratory and urogenital tracts — it seemed improbable that innate immune recognition of microorganisms could be coupled to the immediate initiation of immune responses against them without leading to overt, organism-wide inflammation and its damaging effects. It was therefore hypothesized that microbial sensing at the body surface needs to be tightly controlled to ensure a symbiotic relationship between the host and its indigenous commensal microorganisms[3], while allowing for the initiation of a rapid, sterilizing immune response on penetration of microorganisms into non-colonized sites. This idea was developed further after the realization that host–microbiota mutualism is lost in the absence of innate immune recognition of commensal microorganisms, with detrimental consequences for health[4,5]. The crosstalk between innate immunity and the microbiome is now known to extend far beyond the achievement of a careful balance between tolerance to commensal microorganisms and immunity to pathogens. The microbiota integrates into whole-organism physiology and influences multiple facets of organismal homeostasis through its effects on the innate immune system. Sensing by this system therefore serves as a rheostat for the metabolic activity of the microbiota and its exposure to diet and xenobiotics, as well as for the presence of mucosal infections. The information that is gathered is then processed at various levels of physiology to dynamically adjust the activity of the host to fit the state of the surrounding microbial ecosystem. Conversely, the innate immune system plays an important part in shaping the community and ecology of indigenous microorganisms into configurations that can be tolerated by the host and are beneficial for its metabolic activities. This complex, bilateral interaction between the host and its microbiota has a crucial role in human health. Many 'multifactorial' disorders, formerly considered to be idiopathic, might therefore be influenced or even driven by alteration of the intimate crosstalk that occurs between the innate immune system and the microbiota during homeostasis. In this Review, we highlight paradigms of interactions between the innate immune system and the microbiota, the mechanisms that are involved in this crosstalk and how aberrations in either of the partners of this communication network contribute to the molecular aetiology of common multifactorial disorders. Because the roles of viruses, fungi and parasites have been summarized elsewhere[6,7], we focus on the interplay between the innate immune system and the bacterial microbiome.

Physiological functions

A network of interactions characterizes the interdependence between the innate immune system and the microbiota[8]. The two systems affect one another to orchestrate whole-organism physiology.

Epithelial cells

Although not classically considered to be bona fide cells of the innate immune system, intestinal epithelial cells are equipped with an extensive repertoire of innate immune receptors[9] (Fig. 1). Expression of these receptors and active signal transduction on microbial recognition is pivotal for intestinal homeostasis because their epithelial-specific deletion leads to breaches in the epithelial barrier, which compromises the spatial separation between commensal bacteria and the lamina propria of the intestines, thereby predisposing the tissue to spontaneous inflammation. This has been demonstrated for components that are involved in TLR signalling, including myeloid differentiation primary response protein MyD88, TNF receptor-associated factor 6 (TRAF6), and NF-κB essential regulator (NEMO)[4,10–12], as well as for orchestrators of cell death such as receptor-interacting serine/threonine-protein kinase 1 (RIPK1), FAS-associated death domain protein (FADD) and caspase-8 (refs 13–16).

NOD-containing protein 2 (NOD2), which is highly expressed in the Paneth cells of the small intestine, is activated by microbial

[1]Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Division of Internal Medicine, Tel Aviv Sourasky Medical Center, Tel Aviv 64239, Israel.[3]Research Center for Digestive Tract and Liver Diseases, Tel Aviv Sourasky Medical Center, , Tel Aviv 64239, Israel. *These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved.

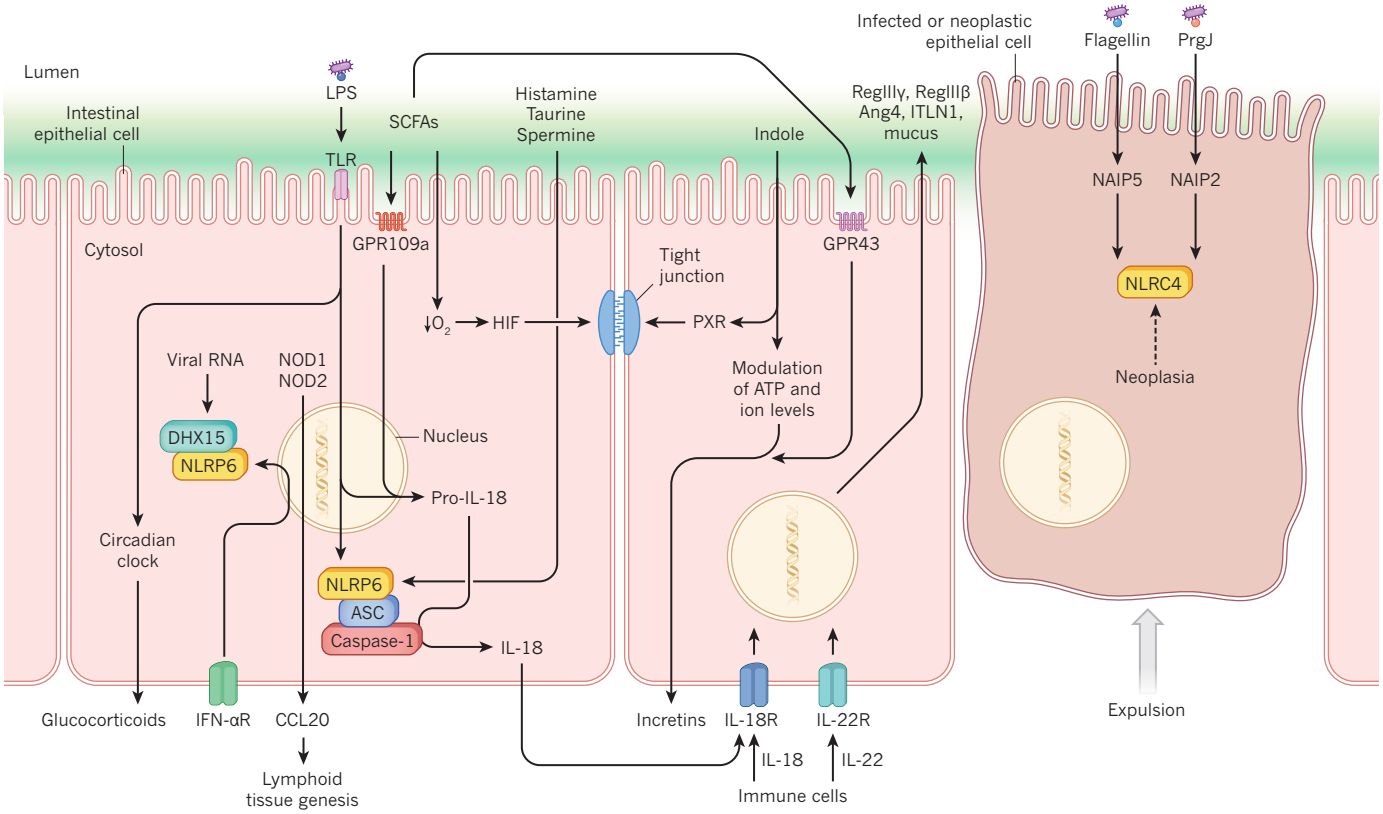


Figure 1 | Intestinal epithelial cells orchestrate the host–microbiota interface. Intestinal epithelial cells use the recognition of microbial-cell components and metabolites to adjust their antimicrobial programme and metabolic homeostasis. The activation of PRRs, such as TLRs and the NOD-like receptors NOD1 and NOD2, is directly coupled to the production of antimicrobial peptides (including RegIIIγ, RegIIIβ, Ang4 and Itln1) and of mucus. IL-18 plays an important part in this process through an autocrine loop. The secretion of epithelial IL-18 requires transcriptional activation through TLRs or the G-protein-coupled receptor GPR109a and posttranscriptional cleavage through the NLRP6 inflammasome. NLRP6 can also be induced by type I interferons and functions as a sensor of viral DNA with pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15. IL-18 and IL-22 derived from immune cells also help to regulate the antimicrobial responses of epithelial cells. CCL20, which is derived from epithelial cells downstream of NOD1 signalling, is involved in the genesis of lymphoid tissue. NLRC4 promotes the expulsion of neoplastic or infected cells from the intestinal epithelium. PRR signalling also orchestrates the circadian clock within intestinal epithelial cells and adjusts the secretion of epithelial-derived metabolic hormones, such as glucocorticoids. Epithelial cells also respond to the levels of microbiota-modulated metabolites, such as SCFAs (acetate, butyrate, propionate), polyamines (spermine), as well as amino acids and products that are derived from them (taurine, histamine, indole). Taurine, histamine and spermine modulate the activity of inflammasome component NLRP6. Indole modulates the levels of incretin section and promotes the barrier function of the epithelium through the PXR, which helps to fortify tight junctions between cells. SCFAs serve as energy sources for epithelial cells and also support barrier function through HIF. ASC, apoptosis-associated speck-like protein containing a CARD; R, receptor.

peptidoglycan and generates a cellular response that includes the secretion of cytokines, the induction of autophagy, intracellular vesicle trafficking, epithelial regeneration and the production of antimicrobial peptides, thereby influencing the composition of the microbiota[17–19]. Epithelial NOD1 is important for both the C–C motif chemokine 20 (CCL20)-mediated generation of isolated lymphoid follicles in the intestine and homeostatic bacterial colonization[20].

PRRs in the epithelium are also important for the elimination of pathogenic infection. Epithelial expression of the inflammasome-forming NLR family CARD-domain-containing protein 4 (NLRC4), a sensor of flagellin and bacterial secretion systems, promotes the expulsion of infected intestinal epithelial cells, thereby contributing to the elimination of enteric pathogens[21,22]. NLRC4 also protects the host from intestinal carcinogenesis[23,24], which provides evidence for a unified model in which epithelial NLRC4 protects the epithelial layer by identifying and dislodging cells that have undergone harmful insults.

Signalling by the NACHT-, LRR- and PYD-domain-containing protein NLRP6 in intestinal epithelial cells is modulated by levels of amino acids and polyamines in the lumen of the intestine. It regulates the interface between the host and microorganisms through the production of inflammasome-mediated interleukin (IL)-18 and the downstream expression of antimicrobial peptides[25], and it also controls the secretion of mucus by goblet cells[26]. Deficiency in NLRP6 leads to imbalances in the composition and function of the microbiota (dysbiosis), altered microbial biogeography and enhanced susceptibility to enteric infection[25–28]. Furthermore, NLRP6 has been described as a regulator of intestinal antiviral immunity[29], which suggests that it might function in the control of both bacterial and viral parts of the microbiome.

Other receptors also integrate microbial signals to adjust IL-18 levels, including hydroxycarboxylic acid receptor 2 (or G-protein-coupled receptor 109A), which is a receptor for butyrate and niacin[30,31], the DNA sensor interferon-inducible protein AIM2 (ref. 32) and the inflammasome component NLRP3. As a consequence, genetic deletion of these receptors leads to intestinal inflammation, tumorigenesis and susceptibility to enteric infection[33,34], which underlines the central role for epithelial IL-18 in orchestrating the intestinal host–microbial interface.

Intriguingly, the impact of microorganisms on intestinal epithelial cells extends far beyond the classical immunological functions of these cells. Commensal colonization probably has a major role in the metabolism of intestinal epithelial cells. Microbiota-derived short-chain fatty acids (SCFAs) serve as an energy source for the epithelium and they affect both oxygen consumption and hypoxia-inducible factor (HIF)-mediated fortification of the epithelial barrier[35]. The microbial metabolite indole promotes barrier function through the pregnane X receptor (PXR; also known as nuclear receptor subfamily 1 group I member

© 2016 Macmillan Publishers Limited. All rights reserved.

2 (NR1I2))[36] and increases the secretion of glucagon-like peptide-1 (GLP-1), an incretin with profound influences on host metabolism[37]. Microbiota-induced TLR signal transduction in intestinal epithelial cells also drives intestinal hormone production through the coordination of the circadian clock, a transcription-factor network that rhythmically controls the diurnal succession of cellular metabolic activity[38]. The microbiota itself undergoes rhythmic oscillations in composition and function[39,40], which suggests that the varying levels of microbial influence on the innate immune system might underlie marked fluctuations over the course of a day.

Taken together, intestinal epithelial cells integrate microbial signals into both the orchestration of the host–microbial interface, which consists of mucus and antimicrobial peptides, and the dynamic adjustment of cellular metabolism (Fig. 1).

Myeloid cells

Germ-free mice have a profoundly altered innate immune system. The microbiota influences the development and function of myeloid cells in multiple organs and at different time points during cellular development (Fig. 2). In the absence of the microbiota, myeloid-cell development in the bone marrow is reduced, which results in the delayed clearance of systemic bacterial infection[41]. The level of myelopoiesis correlates with the complexity of the intestinal microbiota and is adjusted in accordance with the level of TLR ligands that are present in blood serum[42]. Microbiota-derived SCFAs might similarly drive myelopoiesis in the bone marrow[41,43]. The influence of the microbiota on myelopoiesis begins before birth. The offspring of mice that are treated with antibiotics during pregnancy have lower numbers of blood neutrophils and their bone-marrow precursors[44], and gestational colonization with microorganisms increases the number of intestinal mononuclear cells in newborn mice[44].

The microbiome also influences the maturation of myeloid cells after haematopoiesis. The continuous presence of microbiota-derived TLR ligands drives the ageing of neutrophils[46]. The number of circulating basophils is likewise influenced by microbiome-derived TLR ligands[47].

In addition to affecting circulating myeloid cells, the microbiota strongly influences the biology of tissue-resident macrophages. Microglia, the macrophages of the central nervous system, display an altered morphology in germ-free mice — a phenotype that is, in part, due to a paucity of SCFAs[48]. In the skin, the microbiota influences the composition and inflammatory potential of resident myeloid cells[49]. In the lungs, treatment with antibiotics causes a shift in macrophage polarization that is mediated by prostaglandin E2, which enhances susceptibility to allergic airway inflammation[50]. In the intestine, microbial SCFAs serve as a signal to alter the gene-expression profile of local macrophages[31,51]. The microbiota also regulates the trafficking of myeloid cells in the gut. Intestinal microbial colonization drives the continuous replenishment of macrophages in the intestinal mucosa by monocytes that express C–C chemokine receptor type 2 (CCR2)[52].

The tissue-specific effects of the microbiome on resident myeloid cells go beyond bona fide immunological functions. Signals that are released by the microbiota might influence the interactions between neurons of the enteric nervous system and intestinal muscularis macrophages to facilitate gastrointestinal motility[53]. Commensal microorganisms regulate both the expression of bone morphogenetic protein 2 (BMP2) by muscularis macrophages and the production of colony-stimulating factor 1 (CSF1; also known as macrophage colony-stimulating factor 1) by enteric neurons, which in turn influences smooth-muscle contractions in the intestinal muscle layer[53]. The microbiome also has an influence on tissue recovery after injury. A 2015 study found that the intestinal microbiota sustains inflammation and lymphadenopathy after infection with *Yersinia pseudotuberculosis*[54], thus compromising the return to homeostatic tissue-specific immunity.

Such findings suggest that colonization by commensal microorganisms profoundly shapes the myeloid landscape of the host, both in

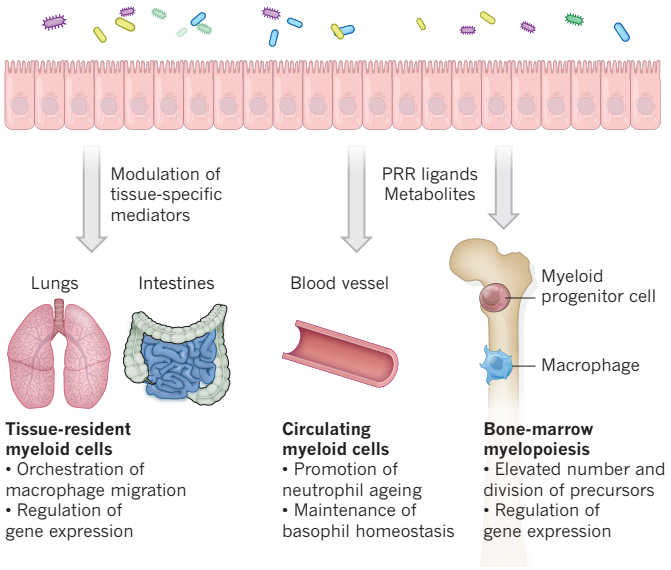


Figure 2 | The integration of microbial signals by myeloid cells. The microbiome influences the function of myeloid cells at all stages of their development. The influence of the microbiome on the migration and gene expression of tissue-resident myeloid cells is achieved mainly through the modulation of local metabolites and mediators of tissue identity. Circulating granulocytes are influenced by microbial PRR ligands. Myelopoiesis in the bone marrow is reduced in the absence of commensal bacteria and their microbial products in the blood.

mucosal tissues and systemically. Local concentrations of microbiota-derived metabolites, as well as systemic levels of microbial products, seem to drive myeloid-cell differentiation and function through PRR signalling. Notably, these microbiota-driven alterations in the myeloid-cell pool greatly influence the susceptibility of the host to a variety of disorders, which range from infection[55] and sepsis[44,46] to allergy, asthma[47,50] and graft-versus-host disease[56]. They also regulate the effectiveness of vaccination[57] and therapies for cancer[58].

Innate lymphoid cells

The influence of the microbiota is not limited to the development of the myeloid arm of the innate immune system. However, the regulation of innate lymphoid cells by the microbiota seems to follow rules and mechanisms that are different from the principles applied to myeloid-cell regulation (Fig. 3). Innate lymphoid cells (ILCs), a recently discovered lymphocyte branch of the innate immune system, develop normally in the absence of the microbiota[59], but the proper functioning of ILCs is dependent on commensal microbial colonization[60–62]. Rather than exerting their effect during lymphopoiesis, signals that stem from commensal microorganisms seem to influence the maturation and acquisition of the tissue-specific functions of ILCs.

The ILC family consists of cytotoxic cells (natural killer cells) and non-cytotoxic subsets (ILC1, ILC2 and ILC3). Most studies that examine the influence of the microbiota on ILCs have focused on ILC3. The importance of ILC3 cells in host–microbiota interactions became clear when their depletion — and the resulting abrogation of IL-22 production — was shown to produce a loss of bacterial containment in the intestine[63]. The microbiota also influences ILC3 interactions with other components of the immune system. The presentation of microbial antigens by ILC3s limits commensal-specific T-cell responses[64] to maintain tolerance to commensal bacteria[65]. Microbial sensing and the production of IL-1β by intestinal macrophages drive granulocyte–macrophage colony-stimulating factor (GM-CSF) secretion by ILC3s, which is required for macrophage function and the induction of oral tolerance[66]. Flagellin sensing by myeloid cells that carry the CD103 antigen is required for the IL-23-mediated production of IL-22 by ILCs[67]. Furthermore, the production of lymphotoxin-α (also known as tumour necrosis

© 2016 Macmillan Publishers Limited. All rights reserved.

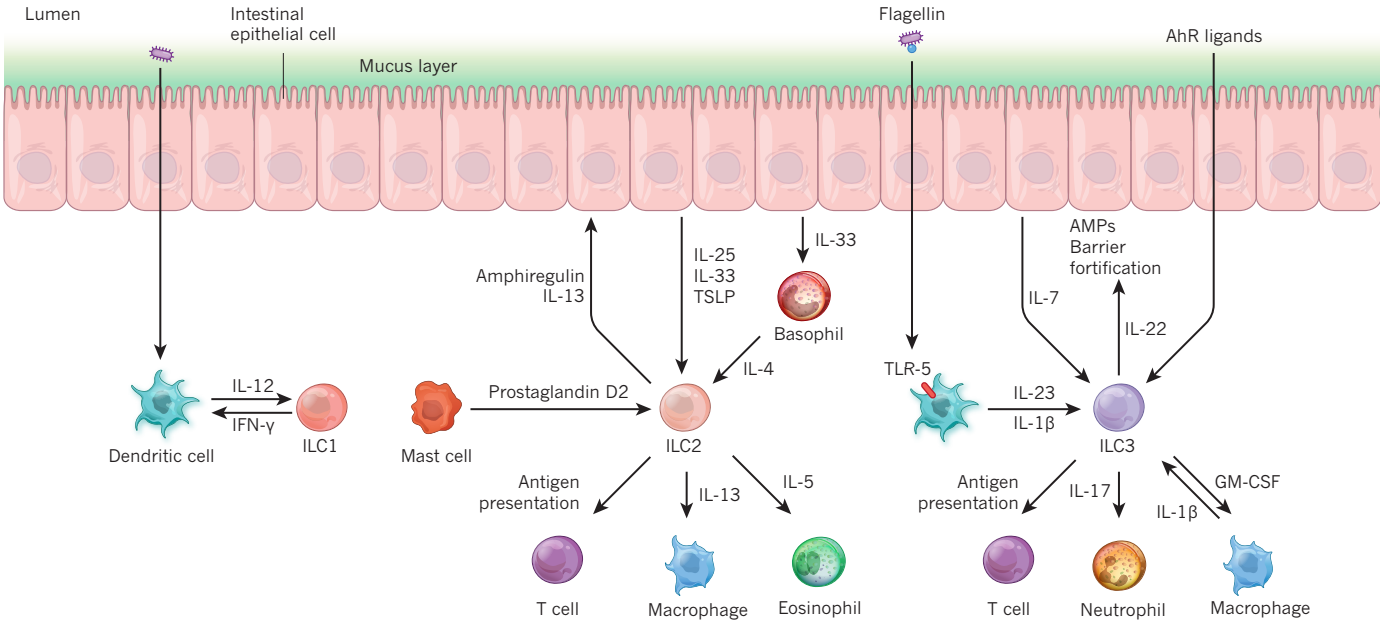


Figure 3 | The integration of microbial signals by ILCs. ILCs communicate with the local microbiota through cytokines, PRR ligands and antimicrobial peptides. In many cases, epithelial cells or myeloid cells serve as relay stations for crosstalk between ILCs and the microbiota. Group 1 ILC (ILC1) cells can be activated by myeloid-cell-derived IL-12. Group 2 ILC (ILC2) cells are activated by epithelial-derived cytokines and orchestrate type 2 immunity through their interactions with mast cells, eosinophils, basophils and macrophages. Group 3 ILC (ILC3) cells interact with cells of both the innate and adaptive immune systems. They also secrete IL-22, which initiates an antimicrobial programme as well as barrier fortification in epithelial cells. AhR, aryl hydrocarbon receptor; AMPs, antimicrobial proteins; LT, lymphotoxin; TSLP, thymic stromal lymphopoietin.

factor-β) by ILC3s is crucial for the production of IgA and for microbiota homeostasis in the intestine[68]. An equally important microbiota-instructed function of ILCs is their communication with epithelial cells. Microbiota-induced IL-22 production by ILC3s induces expression of the enzyme fucosyltransferase 2 (galactoside 2-α-L-fucosyltransferase 2) and fucosylation of surface proteins by intestinal epithelial cells, which is required for host defence against enteric pathogens[69].

Although these examples highlight the importance of microbial signals for the maturation and function of ILCs, the precise mechanisms through which they exert their influence remain unclear and are, in some cases, controversial. For instance, some studies have reported elevated levels of IL-22 by ILCs in the absence of the microbiota, whereas others have documented the abrogation of IL-22 secretion[70]. Different conclusions have also been reached in relation to whether the number of tissue-resident ILCs is altered in mice that are germ-free or have been treated with antibiotics[70]. Further studies are needed to reconcile these observations and their underlying mechanisms.

The microbiota might also influence the activity of the other ILC subsets. ILC2s are activated by epithelial tuft-cell-derived IL-25 (ref. 71), which is produced in a microbiota-dependent manner[62]. Deletion of the ILC1-lineage transcription factor T-bet (also known as T-box transcription factor TBX21) in the innate immune system results in ILC-dependent and *Helicobacter typhlonius*-driven inflammation of the intestines[72].

Collectively, the myeloid and lymphoid branches of the innate immune system are shaped by the microbiota, but the underlying mechanisms are based on distinct principles. A scenario could be envisioned in which the complexity of commensal microbial colonization is reflected in the amount of circulating PRR ligands and the concentrations of microbiota-derived metabolites in tissues, both of which tune the level of myelopoiesis, as well as the system's inflammatory capacity, over the short-term. By contrast, ILC development might be hardwired to anticipate microbial colonization. Tissue-resident ILCs would then integrate signals from the microbiota, through regulatory mechanisms that are not fully understood, to fine-tune innate and adaptive immune responses at the tissue level.

Effects of the innate immune system on the microbiome

On sensing information about the metabolic state of the microbiota, the innate immune system relays signals to the host to adapt tissue-level physiology and might also adjust the composition and function of the microbiota. Genetic evidence from humans and mice indicates that the innate immune system plays an important part in regulating variations in microbiota composition over time and between individuals[73]. Dysbiosis has been reported in several mouse models of innate immune deficiency[8], such as in mice that lack the genes *NOD2* (refs 17, 19, 74), *NLRP6* (ref. 27) or *TLR5* (ref. 75). The innate immune system might therefore function to promote the growth of beneficial members of the microbiota and to contribute to the maintenance of a stable community of microorganisms. This is best demonstrated by the induction of epithelial fucosylation by ILC3s and IL-22. During starvation that is associated with intestinal infection, the shedding of fucosylated proteins into the intestinal lumen serves as a source of energy for commensal bacteria[76]. Innate-immune-system resources therefore can be mobilized to support the microbiota during perturbations of the intestinal ecosystem. Similarly, TLR1 signalling is required to maintain the composition of the microbiota after *Yersinia enterocolitica* infection[77]. By contrast, PRRs do not seem to play a part in the development of the microbiota after treatment with antibiotics has ended[78]. However, it remains possible that activities of the microbiota that are independent of PRRs are involved in controlling the succession of microbial colonization after catastrophic events in the ecosystem.

The mechanisms through which the microbiota controls the development of the innate immune system are beginning to be understood, although the principles and purpose of innate-immune control over temporal dynamics in microbiota function remain unknown. Future mechanistic studies need to better define the characteristics of a 'healthy' microbiome that the host immune system attempts to preserve. Insights into such mechanisms came from the finding that dysbiosis in NLRP6-deficient mice was associated with similar metagenomic functions as were being studied in different animal facilities[25]. Dysbiosis developed *de novo* after the colonization of germ-free NLRP6-deficient mice, which indicates that certain PRRs might create specific

© 2016 Macmillan Publishers Limited. All rights reserved.

antimicrobial landscapes that are associated with the preservation of distinct functions of the microbiome.

Mechanisms of system crosstalk

A wide range of physiological contexts are influenced by communication between the microbiota and the innate immune system, and it is interesting to consider the molecular and cellular mechanisms that mediate this communication at the functional level. Commensal microbial colonization is known to influence the activity of the innate immune system according to a number of common principles.

Transcriptional reprogramming

One of the most striking observations made in germ-free mice was the reprogramming of intestinal gene expression in animals that were colonized with a single commensal bacterium[79] or a single enteric virus[80]. This includes the expression of genes that are involved in host nutrient absorption and processing, barrier functions, gut motility, intestinal immune responses, angiogenesis and the metabolism of xenobiotics. Studies of germ-free mice and of natural microbial colonization during postnatal development have substantiated such findings by showing that transcriptional reprogramming of the intestine by the microbiota spans different regions of the gastrointestinal tract and is partially dependent on microbial sensing receptors of the innate immune system[81,82]. The impact of the microbiome on transcription reaches beyond the intestine. For instance, the livers of germ-free mice show massive alterations in the expression of a range of genes with metabolic and non-metabolic functions[83].

The transcriptional responses of the host to bacterial colonization are in part evolutionarily conserved, as shown by reciprocal microbiota transplantations between mice and zebrafish[84]. Yet there is a considerable degree of species specificity in host responses to microbial colonization, especially with respect to the maturation of the immune system[85]. Although such examples underline the importance of transcriptional responses to commensal colonization for the innate immune system, several lines of evidence suggest that regulation also occurs through mechanisms other than gene expression. Constituents of the microbiota have been implied in the regulation of ubiquitin signalling[86], protein neddylation[87,88], the nuclear translocation of RelA (also known as transcription factor p65) (ref. 89) and vesicle trafficking[90], which indicates that the full regulatory reach of commensal microorganisms is yet to be defined.

Epigenetic programming

Because a large fraction of the transcriptome is shaped by the microbiome in an organ-specific manner, gene regulatory mechanisms must integrate microbial signals into the orchestration of gene expression. Although it is appreciated that bacterial pathogens can modulate host epigenomics, the epigenetic interpretation of commensal microbial colonization by the innate immune system is only starting to be investigated. On an organismal scale, mediation of the transcriptional reprogramming of gene expression in the intestine by the open chromatin landscape was ruled out because the chromatin accessibility in germ-free mice is similar to that in colonized mice[91]. Instead, microbial regulation of gene transcription in the host might be achieved by differential expression of specific transcription factors and their binding to chromatin. The exploration of this possibility on an organismal scale could reveal potential regulatory pathways through which information on the state of the microbiota is integrated into the chromatin landscape of host tissues.

Specific examples of this phenomenon exist in the context of the innate immune system. Analysis of epigenetic modifications in the intestinal epithelial cells of germ-free mice revealed a low level of methylation on the gene that encodes the lipopolysaccharide sensor TLR4, which indicates that commensal bacteria might induce tolerance through the epigenetic repression of PRRs[92]. Microbial colonization of germ-free neonatal mice was found to decrease the methylation level of the chemokine-encoding gene *Cxcl16*, which reduced its expression and diminished the recruitment of invariant natural killer T cells, ameliorating colitis and allergic asthma[93]. A comparison of mononuclear phagocytes from colonized and germ-free mice revealed that the microbiota promotes the trimethylation of histone H3 at lysine 4 at the loci of inflammatory genes, including those which encode the type I interferons[55]. The acetylation of histones is similarly involved in the crosstalk between the microbiota and the innate arm of the immune system. When histone deacetylase 3 is specifically deleted from intestinal epithelial cells, gene expression is massively altered and the integrity of the epithelial barrier is lost[94]. These aberrations are known to be microbiota-dependent because germ-free mice that lacked intestinal histone deacetylase 3 do not present the same phenotype as their colonized counterparts[94].

Although the microbial signals that are responsible for specific epigenetic alterations are mostly unknown, it seems probable that microbial metabolites, rather than just the presence or absence of microorganisms, mechanistically influence the orchestration of histone modifications. For instance, the microbiota-derived SCFA butyrate was shown to modulate the immune response of colonic macrophages through the inhibition of histone deacetylases[51], with a potential contribution to the maintenance of immunological tolerance to commensal microorganisms. Transcriptional reprogramming through epigenetic modifications is therefore a prominent mechanism by which the microbiota exerts its influence on host innate immunity. The elucidation of the precise mechanisms through which microbial molecules influence host-cell epigenomes and adjust the transcriptome to respond to the state of microbial colonization is an exciting area for future research.

Hierarchical feedback loops

The local containment and functional maintenance of a microbial ecosystem within the host is a formidable challenge for the mammalian innate immune system. Co-evolution between the microbiota and

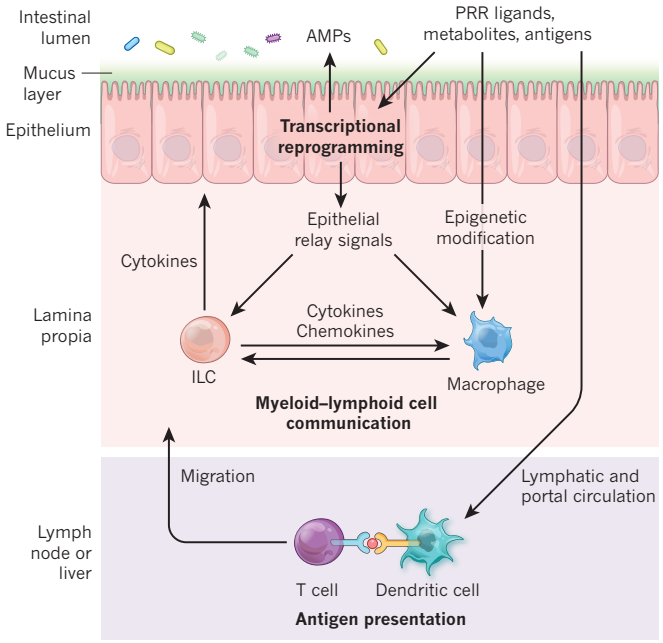


Figure 4 | The hierarchy of anatomy in microbiome–innate-immune-system interactions. Feedback loops between the host and the microbiome can be restricted to the epithelial layer of the intestinal wall, in which they consist of a brief circuit that links microbial sensing with transcriptional reprogramming and antimicrobial responses. A prototypical cytokine for such communication is the paracrine IL-18. Feedback loops that extend to the underlying lamina propria involve communication between epithelial, myeloid and lymphoid cells using cytokines and chemokines. Examples of cytokines that mediate such interactions are IL-22 and IL-23. Microbial products can also reach the draining lymph node and liver, where dendritic cells regulate anticommensal T-cell immunity to promote microbial containment. AMPs, antimicrobial peptides.

© 2016 Macmillan Publishers Limited. All rights reserved.

the host has led to the development of sophisticated feedback loops to accomplish this task. These loops can be regulated by various layers of cells within the intestinal wall. Although they are often restricted to the epithelium, which is directly exposed to the microbiota, they sometimes extend into the underlying mucosal lamina propria or even the lymphatic and portal circulation (Fig. 4).

In evolutionary terms, feedback loops that are restricted to the epithelium could represent the most ancient form of host–microbiota interaction. Such loops consist of only three steps: first, the recognition of microbes by PRRs; second, the transcriptional response of the host; and third, the secretion of effector molecules. The advantage of using such confined regulatory circuits is that the inflammatory response can be limited to the epithelial layer, without involving entire tissues or multiple organs. Examples include the epithelial-autonomous regulation of antimicrobial-peptide and mucus secretion by NLRP6 and NOD2, as well as the control of intestinal epithelial cell death by NLRC4, which all occur without the apparent contribution of other regulatory layers of cells[17–19,21,22,25–28,74,95–97].

The crosstalk between the innate immune system and the microbiome can also extend to the lamina propria. Microbial sensing by myeloid cells of the lamina propria provides regulatory signals that are crucial for the maintenance of commensal mutualism and the initiation of inflammatory responses in the host[67,98]. Myeloid cells modulate important pathways such as IL-22 production by ILCs, which induces the production of epithelial regenerating islet-derived protein 3 (RegIII)β and RegIIIγ, antimicrobial peptides that are important for maintaining a spatial separation between the majority of commensal bacteria and the intestinal epithelial layer, and this modulation is also pivotal for the local containment of commensals[11,63,99].

Regulatory circuits that reach the lymphatic and portal circulation represent a further level of interaction between the microbiome and the immune system. Migration to the mesenteric lymph nodes of antigen-presenting cells that carry material from commensal gut microbes is essential for the induction of commensal-specific adaptive immune responses[100–102]. Likewise, dendritic cells carry microbial antigens from colonized skin to the draining lymph nodes, where the production of cytokines determines the signature of the anticommensal immune response[103]. A similar 'firewall' might apply in the liver, which microbial products access through the portal vein[104].

Multiple levels of anatomy therefore contribute to the innate immune-mediated containment of the microbiota and to the tailoring of the immune response to the tissue-specific characteristics of host–microbiota interactions.

Impact on diseases

The interactions between the host and its microbiota are crucial for the preservation of tissue homeostasis. It is unsurprising therefore that perturbed interactions have emerged as a pivotal driver of various chronic disease states (see page 94). Three concurrent themes of interactions between the microbiome and the innate immune system are emerging as important contributors to microbiome-mediated disease phenotypes. First, microbial products might serve as perpetual stimuli of chronic immune responses, which contribute to the occurrence of non-resolving inflammation. For instance, microbial signals can sustain inflammation and tissue damage after infection-induced injury to the mucosa[54]. Second, abnormal microbial development during maturation of the innate immune system results in a failure to induce immunological tolerance, which then leads to exacerbated autoimmune and autoinflammatory disorders later in life. An example of this is the condition allergen-induced airway hyperreactivity[105]. Third, the microbiome greatly influences the factors that control tissue-specific immunity through mechanisms that can be active even at sites that are distant from the microbiome[106]. Therefore, dysbiosis can trigger pathophysiologies at remote organs and manifest as distinct symptoms in the context of 'sterile' tissues. For instance, intestinal dysbiosis drives the remodelling of the haematopoietic stem-cell niche in the bone marrow, and it also alters the differentiation of progenitor cells in the context of obesity[107].

A number of medical conditions that occur in people, or the equivalent conditions in animals, demonstrate how aberrations in the crosstalk between the innate immune system and the microbiome can contribute to pathogenesis on a molecular and cellular level (Fig. 5).

Infection

The microbiota contributes to the health of the host by colonizing the mucosal entry sites of pathogens, where it occupies biological niches and prevents invasion of the ecosystem by foreign elements — a concept known as colonization resistance (see page 85). In addition to its direct mediation of niche competition, the microbiota mediates resistance to infection indirectly by stimulating the innate immune response.

A prominent example of this is the intestinal immunity to viral infections that occurs when the host response is impaired by antibiotic-mediated depletion of commensal bacteria[108,109]. Effective antiviral innate immunity in the intestine is achieved through the induction of interferon (IFN)-λ and IL-18 or IL-22 (refs 110, 111) pathways, which then cooperate to induce the activation of signal transducer and activator of transcription 1 (STAT1) and antiviral genes[112]. Although IL-18 and IL-22 are induced by commensal bacteria, the expression of IFN-λ is suppressed by the microbiota, which enables efficient viral persistence[113]. Similarly, certain viruses can hijack interactions between bacterial molecules and the innate immune system, such as LPS–TLR4 signalling, to ensure their efficient transmission[114,115].

The microbiome and innate immune system also cooperate in the eradication of bacterial infection. Sometimes, neither innate immunity nor colonization resistance is sufficient to ensure the expulsion of pathogens. Instead, a combination of the two is required, as in the case of cooperation in the host defence against *Citrobacter rodentium*[116,117], a bacterium that can cause disease in mice. However, such combinatorial responses can be subverted by the pathogen. During infection with *Salmonella* Typhimurium, microbiota-induced IL-22 elicits a response that targets commensal bacteria and liberates a colonization niche for the pathogenic bacterium[118]. *Porphyromonas gingivalis*, an oral bacterium that is associated with periodontitis, evades the host by modulating the TLR2 pathway to support a niche for dysbiosis and subsequent inflammation[119].

Autoimmunity and autoinflammation

Inflammatory bowel disease (IBD) is a group of chronic inflammatory disorders of multifactorial aetiology that affect the gastrointestinal tract and extraintestinal organs. These disorders provide models for studying perturbed crosstalk between the microbiota and the innate immune system because they integrate all aspects of mucosal immunology at the interface between microbial colonization and innate-immune-system activation. They also clearly demonstrate how limitations in our mechanistic understanding of this crosstalk hamper the development of treatments for common human disorders. Dysbiosis has a central role in the pathogenesis of IBD, and the introduction of bacteria that are associated with IBD into a murine model of colitis resulted in chronic disease[120], which suggests that immune dysfunction as an adjunct to specific microbial alteration is necessary for the development of IBD. Despite large-scale efforts, however, no particular species or group of commensal or pathogenic microorganisms has been identified as the cause of IBD in humans. Instead, multiple mechanisms at the interface between the innate immune system and the microbiome, such as microbial sensing, the release of reactive oxygen species and antigen processing, were hypothesized to contribute to the molecular pathophysiology of IBD[121]. Genome-wide association studies in humans have found allelic variance in several of the genes that regulate the innate immune system. These include: *NOD2* (refs 122, 123), which is linked to activation of the immune system by peptidoglycans; *ATG16L1* (refs 124, 125), which has a role in autophagy; and *CLEC7A*[126], which is involved in the recognition of fungi by dendritic cells.

Dysbiosis might also promote other extraintestinal inflammatory

© 2016 Macmillan Publishers Limited. All rights reserved.

and autoimmune disorders, although the underlying mechanism is yet to be completely unravelled. Type 1 diabetes is associated with microbiota compositions that are characterized by low diversity and the expansion of distinct groups of bacteria[127]. Non-obese diabetic mice, an animal model for type 1 diabetes, could be phenotypically rescued by the deletion of the gene *Myd88*. However, germ-free, MyD88-deficient non-obese diabetic mice do develop type 1 diabetes, which could be attenuated by faecal transplantation, demonstrating that microbiota–innate-immune-system interactions can modify the disease[128]. Rheumatoid arthritis was found to associate with an overabundance of *Prevotella copri* and a propensity to develop colitis[129]. Such examples suggest that even classic autoimmune diseases might contain an autoinflammatory component that is driven by perturbed communication between the host and the microbiota.

Interactions between the microbiota and the innate immune system also participate in pulmonary and atopic phenomena. Commensal bacteria have been shown to protect against food allergy and allergic airway inflammation; germ-free mice and mice treated with antibiotics develop exacerbated disease[47,50,130]. Mice that are deficient in TLR2 or TLR4 develop pulmonary damage on the chronic intake of a high-fat diet. This damage is abrogated in germ-free mice or mice that consume antibiotics, and it can be transmitted to wild-type mice by faecal transplantation[131]. Together, these findings reveal the trialogue that exists between the microbiota, the host and environmental factors and that contributes to common idiopathic diseases.

Metabolic syndrome

Obesity has become a global-health problem; in 2014, approximately 40% of the population worldwide was overweight and 13% was obese, according to the World Health Organization. The association of obesity with other metabolic derangements, such as type 2 diabetes, hypertension, dyslipidaemia and non-alcoholic fatty liver disease, is known as metabolic syndrome. This complex of conditions is highly associated with cardiovascular morbidity and mortality, and it has become the leading cause of death worldwide (see page 56).

Obesity and type 2 diabetes are associated with chronic low-grade inflammation and an increased expression of PRRs in adipose tissue, muscle tissue and in circulating monocytes[132]. Both conditions also trigger dysbiosis, which is consistent with the idea that diet and PRR activation shape the microbial composition of the gut[133]. In mice, certain deficiencies of innate-immune receptors induce metabolic aberrations and dysbiosis, which can be transferred to wild-type mice by faecal transplantation and abrogated by treatment with antibiotics[75,106]. The microbiota, innate immunity and metabolic syndrome are directly linked through the secretion of IL-22 by ILCs, a mechanism that was found to preserve the integrity of the intestinal mucosal barrier, thereby alleviating metabolic disorders[134].

Other constituent conditions of metabolic syndrome, such as hypertension and dyslipidaemia, have also been linked to intestinal bacteria. The bacterial composition of stool samples obtained from people with these conditions feature dysbiosis and reduced taxonomic diversity[135,136]. The pathogenesis of non-alcoholic fatty liver disease is linked to interactions between the microbiota and the innate immune system of the host. Deficiencies in inflammasome components exacerbate non-alcoholic fatty liver disease owing to the induction of colonic inflammation and a subsequent increase in the release of TLR agonists from the gut and their arrival at the liver through the portal circulation[106].

Atherosclerosis, a progressive inflammatory process that is another component disorder of metabolic syndrome, involves the accumulation of lipids and the formation of plaques around arterial walls. This pathology was linked to the intestinal microbiota as a result of several observations.

First, the administration of antibiotics was shown to confer beneficial effects on cardiovascular risk factors in a murine model of atherosclerosis[137]. Second, some of the bacterial species in atherosclerotic plaques are common to both the oral and intestinal microbiota, and the presence or absence of these groups correlate with levels of cholesterol in the blood plasma[138]. Third, metabolomic analysis revealed that trimethylamine *N*-oxide, a phospholipid that is found in red meat and is metabolized exclusively by intestinal microbiota, promotes atherosclerosis and increases the risk of cardiovascular diseases[139,140]. Intriguingly, the targeted inhibition of trimethylamine *N*-oxide attenuates features of atherosclerosis, which paves the way for a microbiota-mediated therapeutic approach to the treatment of cardiovascular diseases[141]. Atherosclerosis is also dependent on the host's innate immunity, because a deficiency in *Myd88*, specific TLRs or components of the inflammasome suppresses the condition in murine models[142].

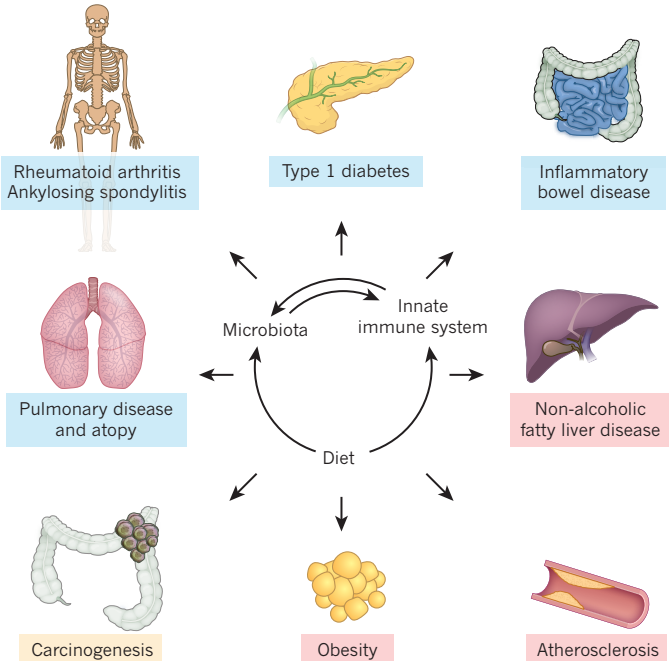


Figure 5 | Microbiome–innate-immune-system interactions are involved in multifactorial diseases. Many inflammatory disorders are influenced by alterations in the crosstalk between innate immunity and the microbiome. These include metabolic (red boxes), neoplastic (orange box) and autoimmune or autoinflammatory (blue boxes) disorders. Modulation of the severity of a disorder through dietary interventions and their influence on microbiome–immune interactions is an exciting area of research.

Cancer

The idea that chronic inflammation drives carcinogenesis has been widely established in various tissues. For example, hepatocellular carcinomas arise in people with chronic hepatitis, colorectal cancer can occur in people with longstanding untreated IBD and Marjolin's ulcers develop on chronically inflamed skin. The presence of bacteria at tumour sites was first described more than a century ago, so it is surprising that the role of the microbiota in tumourigenesis has only recently been recognized. Colorectal carcinogenesis is triggered by a combination of microbiota- and host-dependent mechanisms. Certain bacteria promote carcinogenesis directly, through the secretion of substances that elicit DNA damage[143]. Prominent examples include the excessive release of nitric oxide from immune cells that is triggered by *Helicobacter hepaticus*, the production of reactive oxygen species by *Enterococcus faecalis* and the secretion of an enterotoxin by *Bacteroides fragilis*, which activates the oncogene *c-MYC*. Other bacteria drive carcinogenesis indirectly by sustaining a proinflammatory microenvironment, such as the production by *Fusobacterium nucleatum* of the virulence factor FadA, which increases the paracellular permeability of colonic epithelial cells.

Inflammation might also promote community-level alterations in the microbiome and facilitate bacterial translocation into neoplastic tissue, which further promotes the expression of inflammatory cytokines and leads to the increased growth of tumours[144]. Dysbiosis that arises in

© 2016 Macmillan Publishers Limited. All rights reserved.

the absence of NLRP6 promotes the development of cancer through IL-6-induced epithelial proliferation[17].

The influence of the microbiota on innate immunity has been shown to affect the host response to cancer therapy. For example, germ-free mice and mice that are treated with antibiotics both show a diminished response to immunotherapy by CpG oligonucleotides and chemotherapy owing to the impaired function of myeloid-derived cells in the tumour microenvironment[58]. Furthermore, commensal *Bifidobacterium* enhances immunity to tumours through antibodies directed against programmed cell death 1 ligand 1 (PD-L1) through the augmentation of dendritic-cell function[145]. These studies might open up a fascinating avenue of research to prevent cancer and develop cancer therapeutics through manipulation of the microbiota.

Future directions

The importance of the innate-immune sensing of commensal microorganisms was recognized merely a decade ago. Since then, multiple levels of interaction between the microbiota and the cells of the innate immune system have been uncovered, which range from molecular events at the level of individual cells to the physiology of entire organs. The importance of the microbiome in mammalian health and disease is clearly recognized, and in many cases the innate immune system provides the causal link between disease-associated microbial alterations and the pathophysiological mechanisms of the host. Nonetheless, very few of the insights gained from the study of microbiome–innate-immune-system interactions have been used to develop clinical therapies for inflammatory diseases. In the next decade, research in the field must therefore reach a number of milestones that will help to harness our knowledge to provide clinical applications.

First, the majority of insights so far have been gained from studies of mouse models. The relevance of these principles for microbiome–innate-immune-system interactions in humans remains to be determined.

Second, knowledge of how the microbiome influences the innate immune response is based mostly on well-known examples and might not fully represent the scope of possible mechanisms. Systematic studies that screened members of the microbiome for their effects on the immune system suggest that the range of commensal bacteria that modulate the maturation of the immune system might be far larger than was previously anticipated[146]. Whether the entirety of microbiota–innate-immune-system interactions can be classified according to a limited number of paradigms — that is, whether certain groups of bacteria use common mechanisms to modulate the innate immune system — is still to be uncovered[39].

Third, in comparison to its effect on the adaptive immune system (see page 75), very little is known about the bacterial species, effector molecules and molecular mechanisms through which the microbiota exerts its immune-modulating effect on the cells of the innate arm of the immune system. Because it lacks antigen specificity, the innate immune system might act by broadly evaluating the activity of the microbiome through tissue-level microbial sensing rather than by responding to particular species of bacteria. A comprehensive characterization of the bacterial components and metabolites that are sensed by the innate immune system, through either PRRs or other sensors, as well as their effects on the transcriptional and post-transcriptional landscape of the host, will greatly facilitate our ability to understand the molecular aetiology of microbiome-driven disorders.

Fourth, our deepening knowledge about the interactions between the innate immune system and the microbiome will ultimately result in the development of therapeutic approaches that target these processes. Such interventional strategies, especially when applied to humans, should take into account the enormous variation in both microbiome configurations and innate immune responses that exists between individuals[147]. However, the fact that the microbiome is amenable to rapid change through dietary interventions could be exploited to construct tailored diets that alter microbiome function and downstream innate immune responses to influence common, multifactorial disorders. Dietary modification might alter the microbiome in a way that would enable it to be primed for subsequent immunomodulatory interventions, thereby integrating both treatment modalities (Fig. 5). Alternatively, the identification of 'postbiotic' bioactive microbiome-modulated compounds might allow common downstream pathways in the host to be targeted, thereby influencing the development and outcome of disorders. The future of immunotherapy might therefore combine direct, drug-based immune modulation with microbiome and metabolome modification to collectively target both microbial and host components of the molecular aetiology of disease. ■

Received 8 December 2015; accepted 15 April 2016.

1. Thaiss, C. A., Levy, M., Itav, S. & Elinav, E. Integration of innate immune signaling. *Trends Immunol.* **37,** 84–101 (2016).
2. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308,** 1635–1638 (2005).
3. Shibolet, O. & Podolsky, D. K. TLRs in the gut. IV. Negative regulation of Toll-like receptors and intestinal homeostasis: addition by subtraction. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292,** G1469–G1473 (2007).
4. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by Toll-like receptors is required for intestinal homeostasis. *Cell* **118,** 229–241 (2004).
Refs 4 and 5 highlight the importance of innate-immune-system recognition of the microbiota for host–microbiota homeostasis.
5. Slack, E. *et al.* Innate and adaptive immunity cooperate flexibly to maintain host–microbiota mutualism. *Science* **325,** 617–620 (2009).
6. Pfeiffer, J. K. & Virgin, H. W. Transkingdom control of viral infection and immunity in the mammalian intestine. *Science* **351,** aad5872 (2016).
7. Underhill, D. M. & Pearlman, E. Immune interactions with pathogenic and commensal fungi: a two-way street. *Immunity* **43,** 845–858 (2015).
8. Thaiss, C. A., Levy, M., Suez, J. & Elinav, E. The interplay between the innate immune system and the microbiota. *Curr. Opin. Immunol.* **26,** 41–48 (2014).
9. Pott, J. & Hornef, M. Innate immune signalling at the intestinal epithelium in homeostasis and disease. *EMBO Rep.* **13,** 684–698 (2012).
10. Nenci, A. *et al.* Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature* **446,** 557–561 (2007).
11. Vaishnava, S. *et al.* The antibacterial lectin RegIIIγ promotes the spatial segregation of microbiota and host in the intestine. *Science* **334,** 255–258 (2011).
12. Vlantis, K. *et al.* TLR-independent anti-inflammatory function of intestinal epithelial TRAF6 signalling prevents DSS-induced colitis in mice. *Gut* http://dx.doi.org/10.1136/gutjnl-2014-308323 (2015).
13. Dannappel, M. *et al.* RIPK1 maintains epithelial homeostasis by inhibiting apoptosis and necroptosis. *Nature* **513,** 90–94 (2014).
14. Günther, C. *et al.* Caspase-8 regulates TNF-α-induced epithelial necroptosis and terminal ileitis. *Nature* **477,** 335–339 (2011).
15. Takahashi, N. *et al.* RIPK1 ensures intestinal homeostasis by protecting the epithelium against apoptosis. *Nature* **513,** 95–99 (2014).
16. Welz, P.-S. *et al.* FADD prevents RIP3-mediated epithelial cell necrosis and chronic intestinal inflammation. *Nature* **477,** 330–334 (2011).
17. Couturier-Maillard, A. *et al.* NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. *J. Clin. Invest.* **123,** 700–711 (2013).
18. Nigro, G., Rossi, R., Commere, P. H., Jay, P. & Sansonetti, P. J. The cytosolic bacterial peptidoglycan sensor Nod2 affords stem cell protection and links microbes to gut epithelial regeneration. *Cell Host Microbe* **15,** 792–798 (2014).
19. Ramanan, D., Tang, M. S., Bowcutt, R., Loke, P. & Cadwell, K. Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity* **41,** 311–324 (2014).
20. Bouskra, D. *et al.* Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* **456,** 507–510 (2008).
21. Nordlander, S., Pott, J. & Maloy, K. J. NLRC4 expression in intestinal epithelial cells mediates protection against an enteric pathogen. *Mucosal Immunol.* **7,** 775–785 (2014).
22. Sellin, M. E. *et al.* Epithelium-intrinsic NAIP/NLRC4 inflammasome drives infected enterocyte expulsion to restrict *Salmonella* replication in the intestinal mucosa. *Cell Host Microbe* **16,** 237–248 (2014).
23. Allam, R. *et al.* Epithelial NAIPs protect against colonic tumorigenesis. *J. Exp. Med.* **212,** 369–383 (2015).
24. Hu, B. *et al.* Inflammation-induced tumorigenesis in the colon is regulated by caspase-1 and NLRC4. *Proc. Natl Acad. Sci. USA* **107,** 21635–21640 (2010).
25. Levy, M. *et al.* Microbiota-modulated metabolites shape the intestinal microenvironment by regulating NLRP6 inflammasome signaling. *Cell* **163,** 1428–1443 (2015).
Refs 25 –29 demonstrate the role of epithelial NLRP6 in orchestrating antimicrobial peptide production, mucus secretion and viral recognition.
26. Wlodarska, M. *et al.* NLRP6 inflammasome orchestrates the colonic host–microbial interface by regulating goblet cell mucus secretion. *Cell* **156,** 1045–1059 (2014).
27. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145,** 745–757 (2011).
28. Normand, S. *et al.* Nod-like receptor pyrin domain-containing protein 6

© 2016 Macmillan Publishers Limited. All rights reserved.

(NLRP6) controls epithelial self-renewal and colorectal carcinogenesis upon injury. *Proc. Natl Acad. Sci. USA* **108,** 9601–9606 (2011).
29. Wang, P. *et al.* Nlrp6 regulates intestinal antiviral innate immunity. *Science* **350,** 826–830 (2015).
30. Macia, L. *et al.* Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome. *Nature Commun.* **6,** 6734 (2015).
31. Singh, N. *et al.* Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40,** 128–139 (2014).
32. Hu, S. *et al.* The DNA sensor AIM2 maintains intestinal homeostasis via regulation of epithelial antimicrobial host defense. *Cell Rep.* **13,** 1922–1936 (2015).
33. Man, S. M. *et al.* Critical role for the DNA sensor AIM2 in stem cell proliferation and cancer. *Cell* **162,** 45–58 (2015).
34. Song-Zhao, G. X. *et al.* Nlrp3 activation in the intestinal epithelium protects against a mucosal pathogen. *Mucosal Immunol.* **7,** 763–774 (2014).
35. Kelly, C. J. *et al.* Crosstalk between microbiota-derived short-chain fatty acids and intestinal epithelial HIF augments tissue barrier function. *Cell Host Microbe* **17,** 662–671 (2015).
36. Venkatesh, M. *et al.* Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. *Immunity* **41,** 296–310 (2014).
37. Chimerel, C. *et al.* Bacterial metabolite indole modulates incretin secretion from intestinal enteroendocrine L cells. *Cell Rep.* **9,** 1202–1208, (2014).
38. Mukherji, A., Kobiita, A., Ye, T. & Chambon, P. Homeostasis in intestinal epithelium is orchestrated by the circadian clock and microbiota cues transduced by TLRs. *Cell* **153,** 812–827 (2013).
39. Thaiss, C. A. & Elinav, E. Exploring new horizons in microbiome research. *Cell Host Microbe* **15,** 662–667 (2014).
40. Zarrinpar, A., Chaix, A., Yooseph, S. & Panda, S. Diet and feeding pattern affect the diurnal dynamics of the gut microbiome. *Cell Metab.* **20,** 1006–1017 (2014).
41. Khosravi, A. *et al.* Gut microbiota promote hematopoiesis to control bacterial infection. *Cell Host Microbe* **15,** 374–381 (2014).
42. Balmer, M. L. *et al.* Microbiota-derived compounds drive steady-state granulopoiesis via MyD88/TICAM signaling. *J. Immunol.* **193,** 5273–5283 (2014).
43. Trompette, A. *et al.* Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature Med.* **20,** 159–166 (2014).
44. Deshmukh, H. S. *et al.* The microbiota regulates neutrophil homeostasis and host resistance to *Escherichia coli* K1 sepsis in neonatal mice. *Nature Med.* **20,** 524–530 (2014).
45. Gomez de Agüero, M. *et al.* The maternal microbiota drives early postnatal innate immune development. *Science* **351,** 1296–1302 (2016).
46. Zhang, D. *et al.* Neutrophil ageing is regulated by the microbiome. *Nature* **525,** 528–532 (2015).
47. Hill, D. A. *et al.* Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. *Nature Med.* **18,** 538–546 (2012).
48. Erny, D. *et al.* Host microbiota constantly control maturation and function of microglia in the CNS. *Nature Neurosci.* **18,** 965–977 (2015).
49. Tamoutounour, S. *et al.* Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. *Immunity* **39,** 925–938 (2013).
50. Kim, Y. G. *et al.* Gut dysbiosis promotes M2 macrophage polarization and allergic airway inflammation via fungi-induced PGE2. *Cell Host Microbe* **15,** 95–102 (2014).
51. Chang, P. V., Hao, L., Offermanns, S. & Medzhitov, R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc. Natl Acad. Sci. USA* **111,** 2247–2252 (2014).
52. Bain, C. C. *et al.* Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nature Immunol.* **15,** 929–937 (2014).
53. Muller, P. A. *et al.* Crosstalk between muscularis macrophages and enteric neurons regulates gastrointestinal motility. *Cell* **158,** 300–313 (2014); erratum **158,** 1210 (2014).
54. Fonseca, D. M. *et al.* Microbiota-dependent sequelae of acute infection compromise tissue-specific immunity. *Cell* **163,** 354–366 (2015).
55. Ganal, S. C. *et al.* Priming of natural killer cells by nonmucosal mononuclear phagocytes requires instructive signals from commensal microbiota. *Immunity* **37,** 171–186 (2012).
56. Schwab, L. *et al.* Neutrophil granulocytes recruited upon translocation of intestinal bacteria enhance graft-versus-host disease via tissue damage. *Nature Med.* **20,** 648–654 (2014).
57. Oh, J. Z. *et al.* TLR5-mediated sensing of gut microbiota is necessary for antibody responses to seasonal influenza vaccination. *Immunity* **41,** 478–492 (2014).
58. Iida, N. *et al.* Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* **342,** 967–970 (2013).
59. Sawa, S. *et al.* Lineage relationship analysis of RORγt$^+$ innate lymphoid cells. *Science* **330,** 665–669 (2010).
60. Sanos, S. L. *et al.* RORγt and commensal microflora are required for the differentiation of mucosal interleukin 22-producing NKp46$^+$ cells. *Nature Immunol.* **10,** 83–91 (2009).
61. Satoh-Takayama, N. *et al.* Microbial flora drives interleukin 22 production in intestinal NKp46$^+$ cells that provide innate mucosal immune defense. *Immunity* **29,** 958–970 (2008).
62. Sawa, S. *et al.* RORγt$^+$ innate lymphoid cells regulate intestinal homeostasis by integrating negative signals from the symbiotic microbiota. *Nature Immunol.* **12,** 320–326 (2011).
63. Sonnenberg, G. F. *et al.* Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science* **336,** 1321–1325 (2012).
Refs 63–65 demonstrate a role for innate lymphoid cells in the local containment of the microbiota and in regulating T-cell responses to the microbiota.
64. Hepworth, M. R. *et al.* Innate lymphoid cells regulate CD4$^+$ T-cell responses to intestinal commensal bacteria. *Nature* **498,** 113–117 (2013).
65. Hepworth, M. R. *et al.* Group 3 innate lymphoid cells mediate intestinal selection of commensal bacteria-specific CD4$^+$ T cells. *Science* **348,** 1031–1035 (2015).
66. Mortha, A. *et al.* Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. *Science* **343,** 1249288 (2014).
67. Kinnebrew, M. A. *et al.* Interleukin 23 production by intestinal CD103$^+$CD11b$^+$ dendritic cells in response to bacterial flagellin enhances mucosal innate immune defense. *Immunity* **36,** 276–287 (2012).
68. Kruglov, A. A. *et al.* Nonredundant function of soluble LTα3 produced by innate lymphoid cells in intestinal homeostasis. *Science* **342,** 1243–1246 (2013).
69. Goto, Y. *et al.* Innate lymphoid cells regulate intestinal epithelial cell glycosylation. *Science* **345,** 1254009 (2014).
70. Sonnenberg, G. F. & Artis, D. Innate lymphoid cell interactions with microbiota: implications for intestinal health and disease. *Immunity 37,* 601–610 (2012).
71. von Moltke, J., Ji, M., Liang, H. E. & Locksley, R. M. Tuft-cell-derived IL-25 regulates an intestinal ILC2-epithelial response circuit. *Nature* **529,** 221–225 (2016).
72. Powell, N. *et al.* The transcription factor T-bet regulates intestinal inflammation mediated by interleukin-7 receptor$^+$ innate lymphoid cells. *Immunity* **37,** 674–684 (2012).
73. Levy, M., Thaiss, C. A. & Elinav, E. Metagenomic cross-talk: the regulatory interplay between immunogenomics and the microbiome. *Genome Med.* **7,** 120 (2015).
74. Petnicki-Ocwieja, T. *et al.* Nod2 is required for the regulation of commensal microbiota in the intestine. *Proc. Natl Acad. Sci. USA* **106,** 15813–15818 (2009).
75. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328,** 228–231 (2010).
76. Pickard, J. M. *et al.* Rapid fucosylation of intestinal epithelium sustains host-commensal symbiosis in sickness. *Nature* **514,** 638–641 (2014).
77. Kamdar, K. *et al.* Genetic and metabolic signals during acute enteric bacterial infection alter the microbiota and drive progression to chronic inflammatory disease. *Cell Host Microbe* **19,** 21–31 (2016).
78. Ubeda, C. *et al.* Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J. Exp. Med.* **209,** 1445–1456 (2012).
79. Hooper, L. V. *et al.* Molecular analysis of commensal host-microbial relationships in the intestine. *Science* **291,** 881–884 (2001).
This study provided initial insight into the effects of commensal bacteria on genome-wide transcriptional reprogramming.
80. Kernbauer, E., Ding, Y. & Cadwell, K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature* **516,** 94–98 (2014).
81. Rakoff-Nahoum, S. *et al.* Analysis of gene-environment interactions in postnatal development of the mammalian intestine. *Proc. Natl Acad. Sci. USA* **112,** 1929–1936 (2015).
82. Sommer, F., Nookaew, I., Sommer, N., Fogelstrand, P. & Backhed, F. Site-specific programming of the host epithelial transcriptome by the gut microbiota. *Genome Biol.* **16,** 62 (2015).
83. Leone, V. *et al.* Effects of diurnal variation of gut microbes and high-fat feeding on host circadian clock function and metabolism. *Cell Host Microbe* **17,** 681–689 (2015).
84. Rawls, J. F., Mahowald, M. A., Ley, R. E. & Gordon, J. I. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* **127,** 423–433 (2006).
85. Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149,** 1578–1593 (2012).
86. Patrick, S. *et al.* A unique homologue of the eukaryotic protein-modifier ubiquitin present in the bacterium *Bacteroides fragilis*, a predominant resident of the human gastrointestinal tract. *Microbiology* **157,** 3071–3078 (2011).
87. Neish, A. S. *et al.* Prokaryotic regulation of epithelial responses by inhibition of IκB-α ubiquitination. *Science* **289,** 1560–1563 (2000).
88. Kumar, A. *et al.* Commensal bacteria modulate cullin-dependent signaling via generation of reactive oxygen species. *EMBO J.* **26,** 4457–4466 (2007).
89. Kelly, D. *et al.* Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-γ and RelA. *Nature Immunol.* **5,** 104–112 (2004).
90. Zhang, Q. *et al.* Commensal bacteria direct selective cargo sorting to promote symbiosis. *Nature Immunol.* **16,** 918–926 (2015).
91. Camp, J. G. *et al.* Microbiota modulate transcription in the intestinal epithelium without remodeling the accessible chromatin landscape. *Genome Res.* **24,** 1504–1516 (2014).
92. Takahashi, K. *et al.* Epigenetic control of the host gene by commensal bacteria in large intestinal epithelial cells. *J. Biol. Chem.* **286,** 35755–35762 (2011).
93. Olszak, T. *et al.* Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336,** 489–493 (2012).
94. Alenghat, T. *et al.* Histone deacetylase 3 coordinates commensal-bacteria-dependent intestinal homeostasis. *Nature* **504,** 153–157 (2013).
95. Chen, G. Y., Liu, M., Wang, F., Bertin, J. & Nunez, G. A functional role for Nlrp6

© 2016 Macmillan Publishers Limited. All rights reserved.

in intestinal inflammation and tumorigenesis. *J. Immunol.* **186,** 7187–7194 (2011).
96. Jiang, W. *et al.* Recognition of gut microbiota by NOD2 is essential for the homeostasis of intestinal intraepithelial lymphocytes. *J. Exp. Med.* **210,** 2465–2476 (2013).
97. Kobayashi, K. S. *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* **307,** 731–734 (2005).
98. Franchi, L. *et al.* NLRC4-driven production of IL-1β discriminates between pathogenic and commensal bacteria and promotes host intestinal defense. *Nature Immunol.* **13,** 449–456 (2012).
99. Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Med.* **14,** 282–289 (2008).
100. Diehl, G. E. *et al.* Microbiota restricts trafficking of bacteria to mesenteric lymph nodes by $CX3CR1^{hi}$ cells. *Nature* **494,** 116–120 (2013).
101. Macpherson, A. J. & Uhr, T. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science* **303,** 1662–1665 (2004).
This seminal study defined the lymph-node-restricted 'firewall' circuits that control the local containment of the microbiota.
102. Sano, T. *et al.* An IL-23R/IL-22 circuit regulates epithelial serum amyloid A to promote local effector Th17 responses. *Cell* **163,** 381–393 (2015).
103. Naik, S. *et al.* Commensal–dendritic-cell interaction specifies a unique protective skin immune signature. *Nature* **520,** 104–108 (2015).
104. Balmer, M. L. *et al.* The liver may act as a firewall mediating mutualism between the host and its gut commensal microbiota. *Sci. Transl. Med.* **6,** 237ra66 (2014).
105. Zeissig, S. & Blumberg, R. S. Life at the beginning: perturbation of the microbiota by antibiotics in early life and its role in health and disease. *Nature Immunol.* **15,** 307–310 (2014).
106. Henao-Mejia, J. *et al.* Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482,** 179–185 (2012).
107. Luo, Y. *et al.* Microbiota from obese mice regulate hematopoietic stem cell differentiation by altering the bone niche. *Cell Metab.* **22,** 886–894 (2015).
108. Abt, M. C. *et al.* Commensal bacteria calibrate the activation threshold of innate antiviral immunity. *Immunity* **37,** 158–170 (2012).
109. Ichinohe, T. *et al.* Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc. Natl Acad. Sci. USA* **108,** 5354–5359 (2011).
110. Nice, T. J. *et al.* Interferon-λ cures persistent murine norovirus infection in the absence of adaptive immunity. *Science* **347,** 269–273 (2015).
111. Zhang, B. *et al.* Prevention and cure of rotavirus infection via TLR5/NLRC4-mediated production of IL-22 and IL-18. *Science* **346,** 861–865 (2014).
112. Hernández, P. P. *et al.* Interferon-λ and interleukin 22 act synergistically for the induction of interferon-stimulated genes and control of rotavirus infection. *Nature Immunol.* **16,** 698–707 (2015).
113. Baldridge, M. T. *et al.* Commensal microbes and interferon-λ determine persistence of enteric murine norovirus infection. *Science* **347,** 266–269 (2015).
114. Kane, M. *et al.* Successful transmission of a retrovirus depends on the commensal microbiota. *Science* **334,** 245–249 (2011).
115. Kuss, S. K. *et al.* Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science* **334,** 249–252 (2011).
116. Guo, X. *et al.* Innate lymphoid cells control early colonization resistance against intestinal pathogens through ID2-dependent regulation of the microbiota. *Immunity* **42,** 731–743 (2015).
117. Kamada, N. *et al.* Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science* **336,** 1325–1329 (2012).
118. Behnsen, J. *et al.* The cytokine IL-22 promotes pathogen colonization by suppressing related commensal bacteria. *Immunity* **40,** 262–273 (2014).
119. Maekawa, T. *et al. Porphyromonas gingivalis* manipulates complement and TLR signaling to uncouple bacterial clearance from inflammation and promote dysbiosis. *Cell Host Microbe* **15,** 768–778 (2014).
120. Carvalho, F. A. *et al.* Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice. *Cell Host Microbe* **12,** 139–152 (2012).
121. Rigottier-Gois, L. Dysbiosis in inflammatory bowel diseases: the oxygen hypothesis. *ISME J.* **7,** 1256–1261 (2013).
122. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411,** 599–603 (2001).
123. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411,** 603–606 (2001).
124. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genet.* **39,** 207–211 (2007).
125. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genet.* **39,** 596–604 (2007).
126. Iliev, I. D. *et al.* Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science* **336,** 1314–1317 (2012).
127. Kostic, A. D. *et al.* The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17,** 260–273 (2015).
128. Wen, L. *et al.* Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature* **455,** 1109–1113 (2008).
129. Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2,** e01202 (2013).
130. Stefka, A. T. *et al.* Commensal bacteria protect against food allergen sensitization. *Proc. Natl Acad. Sci. USA* **111,** 13145–13150 (2014).
131. Ji, Y. *et al.* Diet-induced alterations in gut microflora contribute to lethal pulmonary damage in TLR2/TLR4-deficient mice. *Cell Rep.* **8,** 137–149 (2014).
132. Donath, M. Y. & Shoelson, S. E. Type 2 diabetes as an inflammatory disease. *Nature Rev. Immunol.* **11,** 98–107 (2011).
133. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
One of the first studies to link dysbiosis to disease.
134. Wang, X. *et al.* Interleukin-22 alleviates metabolic disorders and restores mucosal immunity in diabetes. *Nature* **514,** 237–241 (2014).
135. Le Chatelier, E. *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* **500,** 541–546 (2013).
136. Yang, T. *et al.* Gut dysbiosis is linked to hypertension. *Hypertension* **65,** 1331–1340 (2015).
137. Rune, I. *et al.* Modulating the gut microbiota improves glucose tolerance, lipoprotein profile and atherosclerotic plaque development in ApoE-deficient mice. *PLoS ONE* **11,** e0146439 (2016).
138. Koren, O. *et al.* Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc. Natl Acad. Sci. USA* **108** (suppl.), 4592–4598 (2011).
139. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Med.* **19,** 576–585 (2013).
Refs 139–141 explore the causative involvement of specific bacterial metabolites in metabolic disease.
140. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472,** 57–63 (2011).
141. Wang, Z. *et al.* Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* **163,** 1585–1595 (2015).
142. Jin, C., Henao-Mejia, J. & Flavell, R. A. Innate immune receptors: key regulators of metabolic disease progression. *Cell Metab.* **17,** 873–882 (2013).
143. Irrazábal, T., Belcheva, A., Girardin, S. E., Martin, A. & Philpott, D. J. The multifaceted role of the intestinal microbiota in colon cancer. *Mol. Cell* **54,** 309–320 (2014).
144. Grivennikov, S. I. *et al.* Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature* **491,** 254–258 (2012).
145. Sivan, A. *et al.* Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350,** 1084–1089 (2015).
146. Ahern, P. P., Faith, J. J. & Gordon, J. I. Mining the human gut microbiota for effector strains that shape the immune system. *Immunity* **40,** 815–823 (2014).
147. Zmora, N., Zeevi, D., Korem, T., Segal, E. & Elinav, E. Taking it personally: personalized utilization of the human microbiome in health and disease. *Cell Host Microbe* **19,** 12–20 (2016).

Acknowledgements We apologize to those authors whose relevant work could not be included owing to space constraints. We thank the members of the Elinav laboratory for discussions. C.A.T. received a Boehringer Ingelheim Fonds PhD fellowship. N.Z. is supported by the Gilead Sciences International Research Scholars Program in Liver Disease. E.E. is supported by: Y. and R. Ungar; the Abisch Frenkel Foundation for the Promotion of Life Sciences; the Gurwin Family Fund for Scientific Research; the Leona M. and Harry B. Helmsley Charitable Trust; the Crown Endowment Fund for Immunological Research; the estate of J. Gitlitz; the estate of L. Hershkovich; the Benoziyo Endowment Fund for the Advancement of Science; the Adelis Foundation; J. L. and V. Schwartz; A. and G. Markovitz; A. and C. Adelson; the French National Center for Scientific Research (CNRS); D. L. Schwarz; the V. R. Schwartz Research Fellow Chair; L. Steinberg; J. N. Halpern; A. Edelheit; grants funded by the European Research Council; a Marie Curie Career Integration Grant; the German–Israeli Foundation for Scientific Research and Development; the Israel Science Foundation; the Minerva Foundation; the Rising Tide Foundation; the Helmholtz Association; and the European Foundation for the Study of Diabetes. E.E. is the incumbent of the Rina Gudinski Career Development Chair.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/28j8r5l. Correspondence should be addressed to E.E. (eran.elinav@weizmann.ac.il).

© 2016 Macmillan Publishers Limited. All rights reserved.

REVIEW

doi:10.1038/nature18848

The microbiota in adaptive immune homeostasis and disease

Kenya Honda[1,2,3] & Dan R. Littman[4,5]

In the mucosa, the immune system's T cells and B cells have position-specific phenotypes and functions that are influenced by the microbiota. These cells play pivotal parts in the maintenance of immune homeostasis by suppressing responses to harmless antigens and by enforcing the integrity of the barrier functions of the gut mucosa. Imbalances in the gut microbiota, known as dysbiosis, can trigger several immune disorders through the activity of T cells that are both near to and distant from the site of their induction. Elucidation of the mechanisms that distinguish between homeostatic and pathogenic microbiota–host interactions could identify therapeutic targets for preventing or modulating inflammatory diseases and for boosting the efficacy of cancer immunotherapy.

Microbiotas that establish mutualistic relationships with their mammalian hosts are able to influence a multitude of physiological functions, often through modulation of the host's immune system. Certain bacteria that inhabit defined niches transmit distinct signals that affect functions of both the innate and adaptive immune systems, which often results in systemic outcomes that are distal to the site of colonization. For example, segmented filamentous bacteria (SFB) induce T helper 17 (T_H17) cells in the small intestine and can trigger autoimmune arthritis in susceptible mice[1,2]. Some species of *Bifidobacterium* can enhance the T-cell-dependent anti-tumour effect of blocking the programmed death 1 (PD-1) pathway[3], and regulatory T (T_{reg}) cells that are induced by bacteria can have systemic anti-inflammatory functions[4,5]. There are only a handful of examples of single species or defined communities of bacteria that can be used to provide insight into the mechanisms by which distinct subsets of lymphocytes are activated and polarized. Efforts to culture and characterize the commensal bacteria of humans and to assess their influence on the host's immune system, which typically involve the colonization of germ-free mice, promise to provide new tools for investigating which cell types and signalling pathways are crucial for the induction of distinct immune responses. The characterization of IgA-coated gut bacteria from mice and humans, which provides a snapshot of the bacteria that are sensed by the cells of the adaptive immune system, has also been a valuable advance. This approach has been used to identify bacteria with potentially colitogenic functions in people with malnutrition[6] and in individuals with inflammatory bowel disease[7] as well as to compare species of bacteria that elicit T-cell-dependent and T-cell-independent IgA-mediated responses in the host[8].

In this Review, we describe progress towards understanding how colonization of the mammalian host by microbes influences the functional diversity and the repertoires of B cells and T cells, with an emphasis on the differentiation of IgA-producing B cells and T cells that carry the CD4 antigen, particularly T_H17 cells and T_{reg} cells that constitute a large proportion of the effector T (T_{eff}) cells of the lamina propria of the intestines. The reciprocal roles of lymphocytes in regulating the microbiota, a topic that has so far received little attention, will also be discussed briefly. It should be noted that insights into the interactions of the microbiota with the immune cells of the host tend to come from studies of mice in controlled environments, which have limited exposure to pathogenic microbes or to the microbiota of wild populations. Housing of laboratory mice together with free-living wild mice results in a constitutive increase in highly differentiated innate and adaptive immune cells, including effector memory T cells that carry the CD8 antigen, in the laboratory mice[9]. The immune profile of these mice matches that of adult humans much more closely than does that of mice kept in specific pathogen-free conditions. The failure of some mouse studies to predict the responses of humans to therapy could therefore be partly because of differences in the microbiotas of the species.

Interactions of the microbiota with B cells and T cells

Studies have suggested roles for diverse species of microbes in regulating the distinct branches of the adaptive immune system. Antigen-specific adaptive immune responses influence the mutualistic relationship between the microbiota and the host, and are mostly directed at the microbes of the gut.

IgA

Mucosal IgA is secreted across the epithelium by binding to the polymeric immunoglobulin receptor, after which it binds to microbes, various components of the diet and to antigens in the lumen of the intestine. IgA coats and agglutinates its targets to prevent their direct interaction with the host[10,11]. This averts potentially harmful stimulation of the immune system in mucous membranes by the contents of the lumen and it also serves to regulate the composition of the microbiota. As well as providing a physical barrier, IgA can control the expression of genes by microbes in the intestine. For example, in the absence of IgA, the commensal bacterium *Bacteroides thetaiotaomicron*, which typically does not trigger inflammation in the human gut, expresses high levels of gene products that are involved in the metabolism of nitric oxide and elicits pro-inflammatory signals in the host[12]. Similarly, mice that are deficient for Toll-like receptor 5 (TLR5) show reduced levels of IgA that is directed against the protein flagellin, which results in aberrant expression of flagella-related genes by a wide range of commensal microbes[13]. IgA that has undergone affinity maturation through somatic hypermutation binds to and selects for particular components of the microbiota, which leads to an increase in the diversity of the microbial community and enhances mutualism between the microbiota and the host[14]. Consistent with this observation, people who are deficient in IgA have more bacteria from taxa with potentially inflammatory properties[15]. Moreover, mice that carry a mutation called

[1]Department of Microbiology and Immunology, Keio University School of Medicine, Shinjuku, Tokyo 160-8582, Japan. [2]RIKEN Center for Integrative Medical Sciences, Tsurumi, Yokohama, Kanagawa 230-0045, Japan. [3]AMED-CREST, Chiyoda, Tokyo 100-0004, Japan. [4]The Helen L. and Martin S. Kimmel Center for Biology and Medicine at the Skirball Institute of Biomolecular Medicine, New York University School of Medicine, New York, New York 10016, USA. [5]The Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA.

© 2016 Macmillan Publishers Limited. All rights reserved.

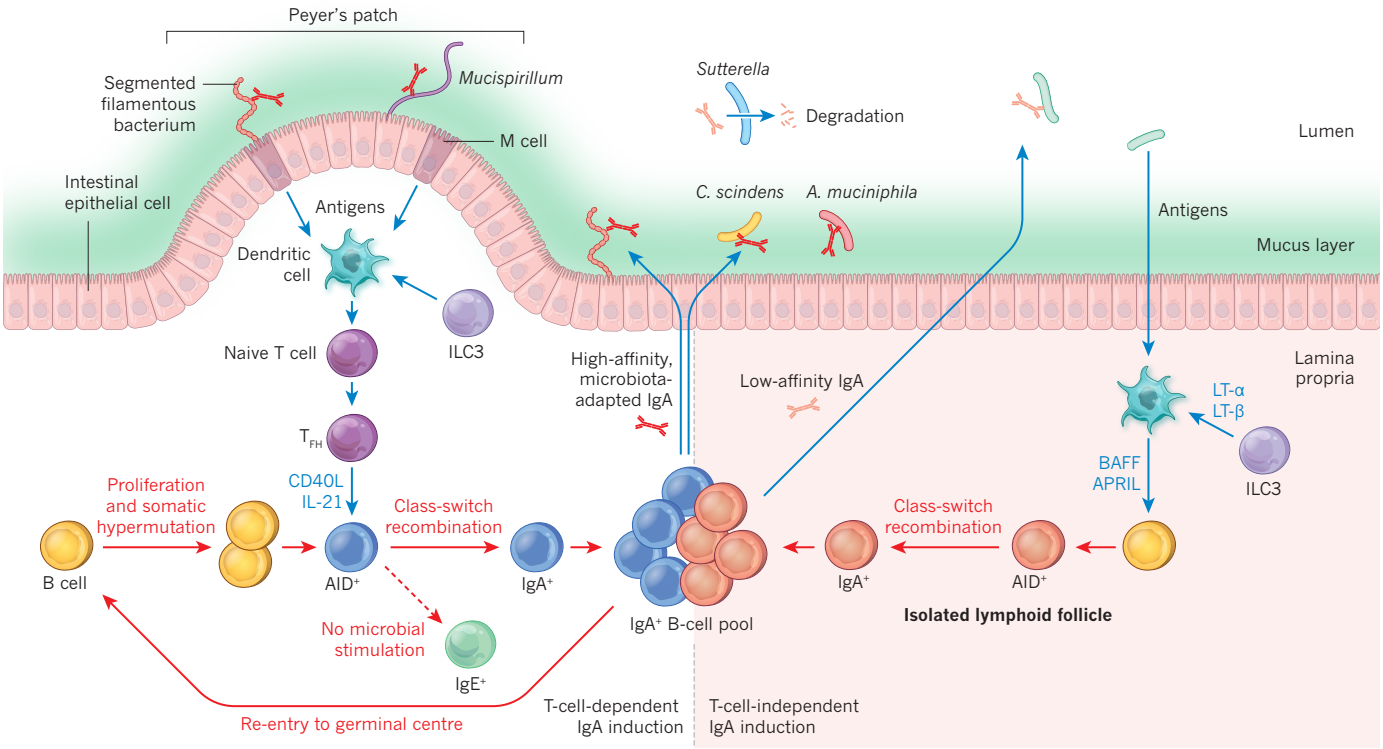


Figure 1 | Induction of IgA in mucosal tissues. T-cell-dependent IgA class-switch recombination (left) takes place mostly in Peyer's patches, in which dendritic cells that are located near to the surface of the epithelium capture antigens from microbes that are transferred by M cells. Dendritic cells induce the differentiation of CD4-expressing T cells into the T follicular helper (T_{FH}) cell subset. CD40 ligand (CD40L) and IL-21 from T_{FH} cells induce the expression of activation-induced cytidine deaminase (AID) in B cells and promote IgA class-switch recombination[11,128]. T-cell-independent IgA class-switch recombination (right) occurs predominantly in the lamina propria and isolated lymphoid follicles (ILFs), where B-cell activating factor (BAFF; also known as TNFSF13B) and its homologue APRIL, which are derived from dendritic cells, promote the induction of AID expression in B cells. Transforming growth factor β (from dendritic cells and stromal cells) and retinoic acid (from dietary vitamin A) play important parts (not shown) in both T-cell-dependent and T-cell-independent pathways. ILC3s that express RORγt also contribute to those pathways, through the expression of lymphotoxin (LT)-α and LT-β, which activate dendritic cells[129]. The gut microbiota affects IgA class-switch recombination in both pathways. The T-cell-independent pathway produces IgA with low affinity but directed towards the microbiota. The T-cell-dependent pathway tends to be activated by bacteria that colonize the surface of the epithelium, such as segmented filamentous bacteria (SFB), *Mucispirillum*, *Clostridium scindens* and *Akkermansia muciniphila*. The IgA-expressing B-cell clones that this pathway induces persist for a long time and can re-enter a germinal centre, where they undergo further somatic hypermutation to produce high-affinity IgA that is adapted to the changing composition of the microbiota.

AID^{G23S} that allows the enzyme activation-induced cytidine deaminase to mediate normal IgA class switching but without somatic hypermutation, harbour a dysbiotic microbiota in their small intestine[16]. Selection of affinity-matured, microbe-specific IgA is therefore crucial for the establishment of a balanced microbiota that, in turn, can restrain inflammatory processes.

Gut plasma cells that produce IgA can be generated by both T-cell-dependent and T-cell-independent mechanisms that involve the cooperation of epithelial cells, dendritic cells and innate lymphoid cells (ILCs) (Fig. 1 and Box 1). In both pathways, the gut microbiota affects the accumulation of cells that express IgA as well as the level and diversity of IgA in the lumen. Indeed, IgA-expressing cells in lymphoid tissue known as Peyer's patches and in the lamina propria are greatly reduced in germ-free animals, and the colonization of germ-free mice with a microbiota quickly triggers the production of IgA. Interestingly, some members of the microbiota, such as species of *Sutterella*, are inversely correlated with the level of IgA in faeces[17]. These members degrade both IgA and a peptide that is required for the stability of IgA in the lumen, known as secretory component. Because microbiota-induced IgAs are directed towards microbial antigens[8], a substantial proportion of the microbiota are coated with IgA and can be detected and characterized through flow cytometry and 16S ribosomal RNA gene sequencing. Known as IgA-SEQ, this combined approach has demonstrated that anatomical location determines whether a particular species of bacterium will elicit an IgA-mediated response in the host[8]. Bacteria that can invade the inner mucous layer of the intestine and colonize regions in proximity to epithelial cells induce high-affinity T-cell-dependent IgA responses[7,8,18]. In particular, SFB and *Mucispirillum* associate intimately with the intestinal epithelium, where they elicit a T-cell-dependent IgA-mediated response and are heavily coated with IgA[8] (Fig. 1). Because SFB have a propensity to induce the production of T_H17 cells, they might also induce follicular helper (T_{FH}) cells with a phenotype that is distinct from those of T_{FH} cells that are induced by other commensal bacteria, thereby resulting in a strong, T_H17-cell-dependent high-affinity IgA response[19]. Mice that are deficient in T cells owing to a lack of T-cell antigen receptor (TCR) chains β and δ, as well as those that lack T_{FH} cells and the T-cell-dependent IgA pathway owing to T-cell-specific inactivation of the gene *Bcl6* in $CD4^+$ T cells, retain an IgA-mediated response that is specific to antigens from commensal bacteria — indicating that the T-cell-independent pathway is also directed at the microbiota[8]. However, this response is characterized largely by the low-affinity binding of IgA to antigens that are shared by multiple species of bacteria[7,11,14].

Induced clones of IgA-producing B cells persist for long periods, even after transient exposure to microbes[20,21] (Fig. 1). Accordingly, an increase in the complexity of the gut microbiota leads to an increase in the diversity of the IgA pool[21]. The repertoire of IgA in the gut is dynamically adjusted in response to changes in the composition of the microbiota[21]. This process of adaptation relies mostly on the re-entry of B-cell clones into a germinal centre and on further somatic hypermutation of B-cell clones that are already established in the pool of plasma cells in the intestine[21]. The types of gut microbes that are targeted by IgA change in accordance with the diet of the host. For example, in mice colonized with the gut

© 2016 Macmillan Publishers Limited. All rights reserved.

microbiotas of undernourished children and fed a nutrient-poor diet, members of the Enterobacteriaceae are heavily coated with IgA[6]. By contrast, in mice that are colonized by the same microbiotas but fed a nutritionally sufficient diet, IgA binds to taxa other than Enterobacteriaceae, even though the load of Enterobacteriaceae is similar. The transfer of Enterobacteriaceae-enriched consortia of IgA-coated microbes leads to a severe enteropathy that is characterized by disruption of the epithelial barrier of the intestine and by weight loss, which suggests that bacteria that are heavily coated with IgA are colitogenic[6]. Consistent with this idea, IgA-coated bacteria that are isolated from people with inflammatory bowel disease promote dramatically exacerbated development of colitis induced by dextran sulfate sodium[7]. However, enteropathy that is induced by colitogenic bacteria can be prevented by the administration of IgA-targeted species of bacteria from healthy microbiotas, such as *Akkermansia muciniphila* and *Clostridium scindens*[6]. Bacteria that are targeted by IgA are therefore not always colitogenic; they can even be of benefit to the host through contributions to enhancing the barrier function of the mucosa.

T_H17 cells

The high-affinity secretory IgA response is proposed to depend largely on T_H17 cells that express RAR-related orphan receptor (ROR)γt[19]. These cells are most abundant in the lamina propria of the intestine, where they account for 30–40% of differentiated memory $CD4^+$ T cells[22–24]. The signature cytokines of T_H17 cells, interleukin (IL)-17A, IL-17F and IL-22, stimulate the production of antimicrobial proteins by intestinal epithelial cells as well as the formation of tight junctions between these cells[25]. They also mediate the transportation of IgA and the recruitment of granulocytes. Consequently, T_H17 cells have an indispensable role in preventing infection by several species of extracellular pathogenic bacteria and fungi. Indeed, genetic defects in the IL-17–IL-17 receptor axis and in RORγt in humans have been linked to susceptibility to chronic mucocutaneous candidiasis[26,27], and a deficiency of both *Il17a* and *Il17f* in mice results in opportunistic infection of mucocutaneous zones by *Staphylococcus aureus*[28]. However, T_H17 cells can also have pathogenic features, particularly following their stimulation with IL-23 and IL-1β[29,30]. Pathogenic T_H17 cells express the pro-inflammatory cytokines interferon (IFN)-γ and granulocyte–macrophage colony-stimulating factor (GM-CSF; also known as CSF2) and exacerbate autoimmune and inflammatory diseases[31–33]. IL-23 is required for the conversion of IL-17-expressing T cells into encephalitogenic and colitogenic T cells that express both IL-17 and IFN-γ or only IFN-γ (known as ex-T_H17 cells, $T_H17.1$ cells or T_H1^* cells)[31,33] and for the onset of disease in mice that are subjected to colitis[34] and to experimental autoimmune encephalomyelitis[29,30]. Although both homeostatic T_H17 cells and pathogenic T_H17 cells are dependent on RORγt in combination with other factors[22,35] for their differentiation, what distinguishes the T_H17 cells that promote homeostatic defence of the gut barrier from those that are involved in pathogenic inflammation is a major unanswered question.

It is unclear whether constituents of the microbiota or other environmental factors direct the differentiation of naive $CD4^+$ T cells into homeostatic or pathogenic T_H17 cells. In experimental models, a multitude of environmental factors have been shown to affect the activation status of intestinal T_H17 cells. For example, a diet that is high in salt enhances the number of T cells in the intestinal lamina propria that express IL-17A and CD4 and increases the risk of T_H17-cell-dependent autoimmunity[36,37]. These phenotypes are ascribed to the salt-mediated induction of serine/threonine-protein kinase Sgk1 (SGK1), which phosphorylates and deactivates forkhead box protein O1, thereby relieving the inhibition of RORγt-mediated transcription of IL-17A and the IL-23 receptor and promoting the generation of pathogenic T_H17 cells[37].

Lipids in the diet have also been implicated in promoting the differentiation of both T_H17 cells and T_{reg} cells[38–40]. Long-chain fatty acids such as lauric acid promote the differentiation of T_H17 cells and induce more severe experimental autoimmune encephalomyelitis, whereas the short-chain fatty acid propionic acid protects animals from disease, in part through the induction of T_{reg} cells[41]. Endogenous fatty acids, which are dependent on the enzyme acetyl-CoA carboxylase 1 for their synthesis, contribute to the differentiation of T_H17 cells and to the development of autoimmune diseases[42]. It has also been suggested that an intermediate in cholesterol biosynthesis acts as an endogenous ligand for RORγt and that enzymes such as CYP51A1 and SC4MOL (also known as MSMO1), which form part of the cholesterol biosynthesis pathway, contribute to T_H17 cell differentiation[43]. These enzymes are upregulated in pathogenic T_H17 cells on their culture with saturated fatty acids, such as palmitic acid, or with IL-23 (ref. 44) (Fig. 2). In the absence of IL-23, non-pathogenic T_H17 cells express the protein CD5L, an inhibitor of fatty-acid synthase, and these cells have elevated levels of polyunsaturated fatty acids at the expense of saturated fatty acids[44]. The mechanism for regulating genes that are the targets of RORγt in the presence of the different types of fatty acids remains unclear, although it is possible that CD5L restricts cholesterol synthesis, which diminishes the endogenous source of RORγt ligands and thus the potential for pathogenicity. Fatty acids that are produced by the microbiota might similarly modulate the activity of RORγt and therefore govern the balance between homeostatic and potentially pathogenic programs of gene expression in T_H17 cells.

The microbiota are the most prominent influence from the environment on the differentiation of T_H17 cells. In germ-free mice, T_H17 cells are scarce in the lamina propria of the intestines as well as in the skin[23,24,45] (Box 2). The number of T_H17 cells in the intestines varies widely between

BOX 1

ILC3s in adaptive immune homeostasis

Signals from the microbiota create complex interactions between epithelial cells, dendritic cells, macrophages and ILC3s. ILC3s contribute to the differentiation of T cells and B cells. For example, ILC3s express lymphotoxin (LT)-α and LT-β and activate dendritic cells, thereby contributing to both T-cell dependent and T-cell-independent pathways of IgA class switching[129]. ILC3s also facilitate the induction of T_H17 cells through the production of IL-22 and other factors. Activation of ILC3s and induction of T_H17 cells have been observed in mice that are colonized by segmented filamentous bacteria (SFB) and other bacteria, including *Citrobacter rodentium*[50,130,131]. Activation of ILC3s by these bacteria requires the TLR-dependent activation of CX_3CR1-expressing cells (derived from monocytes) and their production of IL-23, IL-1β and tumour necrosis factor ligand superfamily member 15 (TNFSF15), which act through receptors on ILC3s[131]. IL-22 from ILC3s then activates epithelial cells to produce serum amyloid A and other factors that are required for the induction of T_H17 cells.

Latent infection of wild-type mice with murine norovirus, which induces pathogenesis in the intestines of mice that lack the gene *Atg16l1* (ref. 132), leads to IL-22 production by ILC3s and the induction of T_H17 cells, while suppressing the expansion of group 2 innate lymphoid cells (ILC2s) — offsetting the deleterious effect of treatment with antibiotics[133]. Viral components of the intestinal microbiota could therefore act with commensal bacteria to reinforce the epithelial barrier through activation of ILC3s and induction of T_H17 cells.

ILC3s that are activated by the microbiota also promote expansion of T_{reg} cells[106]. Gut microbiota induce the production of IL-1β from macrophages in the lamina propria, and this cytokine acts on neighbouring ILC3s to activate their production of CSF2 (ref. 106). CSF2 then acts on CD103-expressing dendritic cells in the colon to enhance the activity of aldehyde dehydrogenase (ALDH) and produce TGF-β and IL-10, which induces T_{reg} cells[106].

© 2016 Macmillan Publishers Limited. All rights reserved.

BOX 2

The skin microbiota and adaptive immunity

The microbiota influences the differentiation of adaptive immune cells both in the skin and the gut. *Staphylococcus epidermidis*, a commensal bacterium of the skin, potently induces T_H17 cells as well as T cells that express both IL-17A and the antigen CD8 (ref. 134). Both cross-presenting dendritic cells that are dependent on basic leucine zipper transcription factor ATF-like 3 (Batf3) and cells derived from monocytes are required to induce a response from cells that express IL-17A and CD8 to *S. epidermidis* in the skin[134]. On infection with the cutaneous pathogenic protozoa *Leishmania major*, local commensal bacteria are necessary to elicit protective immunity (which manifests as inflammation and necrosis), and monoassociation with *S. epidermidis* is sufficient to promote this response[45]. Importantly, T_H17 cells in the skin are affected by the skin microbiota independently of the gut microbiota[45], which suggests that T_H17 cells of the mucosa are regulated in a compartmentalized manner by local commensal bacteria. The production of IL-17A by T cells in the skin requires the expression of IL-1R but not IL-23R, which is in contrast to the requirements of T_H17 cells in the intestines and is consistent with compartment-specific mechanisms for T-cell regulation[45]. Although immunological cross-communication has been shown to occur between mucosal tissues such as the intestine and the lung[135] and the nasopharynx and the uterus[108], there seems to be a compartment-specific regulation of immunity in the skin. This might be because the skin is faced with challenges from the environment that differ from those faced by mucosal sites and therefore requires distinct pathways to control its local immune responses.

animal facilities, even in genetically identical mice that have been reared in specific pathogen-free conditions, and often reflects whether mice have been colonized with SFB[1] (Fig. 2). Such bacteria are potent modulators of the immune-cell functions of the host: as well as inducing T_H17 cells, they also stimulate IgA synthesis[8,46,47] and fucosylation of the epithelium through the activation of group 3 innate lymphoid cells (ILC3s)[48]. SFB that are indigenous to mice and rats are genetically distinct host-specific members of the gut microbiota[49]. On their monocolonization of germ-free mice or rats, populations of SFB can expand in the gut lumen of either species; however, the bacteria bind to epithelial cells of the small intestine and induce T_H17 cells in a strictly host-specific manner[50]. The physical interaction of SFB with the gut epithelium is therefore probably essential for T_H17-cell differentiation. The causality of the relationship between the adhesion of bacteria to the epithelium and the induction of T_H17 cells is further supported by analysis of T_H17-cell induction by the intestinal pathogenic bacteria *Citrobacter rodentium* and *Escherichia coli* O157:H7 (ref. 50). On monocolonization of mice, these species triggered T_H17-cell responses, whereas adhesion-defective mutants fail to do so. Moreover, 20 strains of bacteria that were isolated from the faeces of a person with ulcerative colitis exhibit characteristics that enable their adhesion to epithelial cells and induction of T_H17 cells in the colons of mice[50].

Colonization with adherent SFB elicits a unique program of gene expression that includes the upregulation of two isoforms of the protein serum amyloid A in the epithelial cells of the small intestine. This induction is largely restricted to the terminal ileum, the site at which SFB attach to the epithelium[51]. The genes that encode serum amyloid A are also induced when SFB and epithelial cell lines are cultured together *in vitro*[52], which suggests that their direct interaction initiates a signalling pathway that results in gene expression. In parallel, SFB activate ILC3s to produce IL-22 through the intermediary expression of IL-23 by myeloid cells[51] (Fig. 2). The expression of serum amyloid A in the epithelial cells of the small intestine is dependent on the secretion of IL-22 from ILC3s, by way of phosphorylation of signal transducer and activator of transcription 3 (Stat3) in epithelial cells[50,51]. *In vivo* induction of serum amyloid A might therefore require both adhesion of SFB to epithelial cells and activation of the IL-22 receptor.

Polarization of T_H17 cells that are specific to SFB occurs in the mesenteric lymph nodes, in which RORγt is upregulated before T cells migrate to the lamina propria[51]. T_H17-cell polarization is dependent on monocyte-derived CX_3CR1^+ cells rather than classic dendritic cells[53], although a role for dendritic cells that express CD103 and CD11b and are dependent on Notch2 and IRF4 for their development has also been proposed[54–56]. Polarized T cells that express RORγt and CD4 are distributed broadly throughout the intestine and are even found in the spleen, although most IL-17A expression is confined to the ileum, where serum amyloid A seems to act as an adjuvant and contributes to the induction of IL-17A[51] (Fig. 2).

The mechanism through which serum amyloid A stimulates T_H17 cells has yet to be resolved. In a feed-forward process, myeloid cells including those that carry CX_3CR1 can respond to serum amyloid A by producing cytokines that activate ILC3s, which promotes T_H17-cell differentiation[50] (Box 1). Serum amyloid A might also stimulate T cells directly to enhance RORγt function and upregulate IL-17A expression[51]. Serum amyloid A is a carrier of both high-density lipoprotein and retinol[57], and it can deliver these immunomodulatory molecules to antigen-presenting cells and T cells. The potential regulation of T_H17-cell differentiation by lipids suggests that serum amyloid A might function unconventionally to modulate inflammatory functions in these cells. Together, these findings indicate that the differentiation of T_H17 cells directed by SFB is mediated through a complex circuitry of interactions between epithelial cells, dendritic cells and ILC3s to generate cells that are poised to acquire effector functions in the appropriate microenvironment (Fig. 2). Because SFB have not yet been definitively identified in the human intestine[58], whether this circuitry applies more generally to microbiota-mediated T_H17-cell-induction in humans requires further investigation[58].

Intestinal T_H17 cells and autoimmunity

Most of the T_H17 cells that are elicited by SFB have TCRs that specifically bind to antigens that are expressed by adhesive forms of these bacteria[50]. Two major antigens have been identified as being responsible for this induction[59]. These antigens might be preferentially taken up by the cells of the host when SFB adhere to epithelial cells. Colonization with these bacteria, and the consequent induction of T_H17 cells with TCRs that are specific for SFB antigens, helps to protect the host from intestinal pathogenic species such as *C. rodentium*[1]. However, SFB-induced T_H17 cells might promote pathogenesis in hosts that have a genetic predisposition to autoimmune diseases. In the K/BxN mouse model of autoimmune arthritis, colonization with commensal microbes is required for the development of disease[2]. Monocolonization with SFB enhances the production of autoantibodies and accelerates the progression of disease through the generation of T_H17 cells[2], although a microbiota-induced T_{FH}-cell-dependent process can also precipitate disease[60]. Mice that harbour SFB are more susceptible to experimental autoimmune encephalomyelitis than are germ-free mice[61]. By contrast, the presence of SFB is strongly correlated with a diabetes-free state in non-obese diabetic mice[62]. The influence of such bacteria on the development of autoimmune diseases is therefore dependent on context. The conditions that determine whether intestinal T_H17 cells play a beneficial or harmful part in the host are not yet fully understood. Interestingly, germ-free mice that are colonized with SFB show a striking genotype-specific difference in the induction of T_H17 cells. For instance, BALB/c mice have fewer T_H17 cells but a greater amount and diversity of IgA in their faeces than do C57BL/6 mice[50,63]. Therefore, a combination of genetics and the composition of the gut microbiota affects the status of the immune system and an individual's susceptibility to disease.

In the K/BxN mouse model of autoimmune arthritis, self-reactive T_H17 cells that express a transgenic TCR that is specific for a self antigen can

© 2016 Macmillan Publishers Limited. All rights reserved.

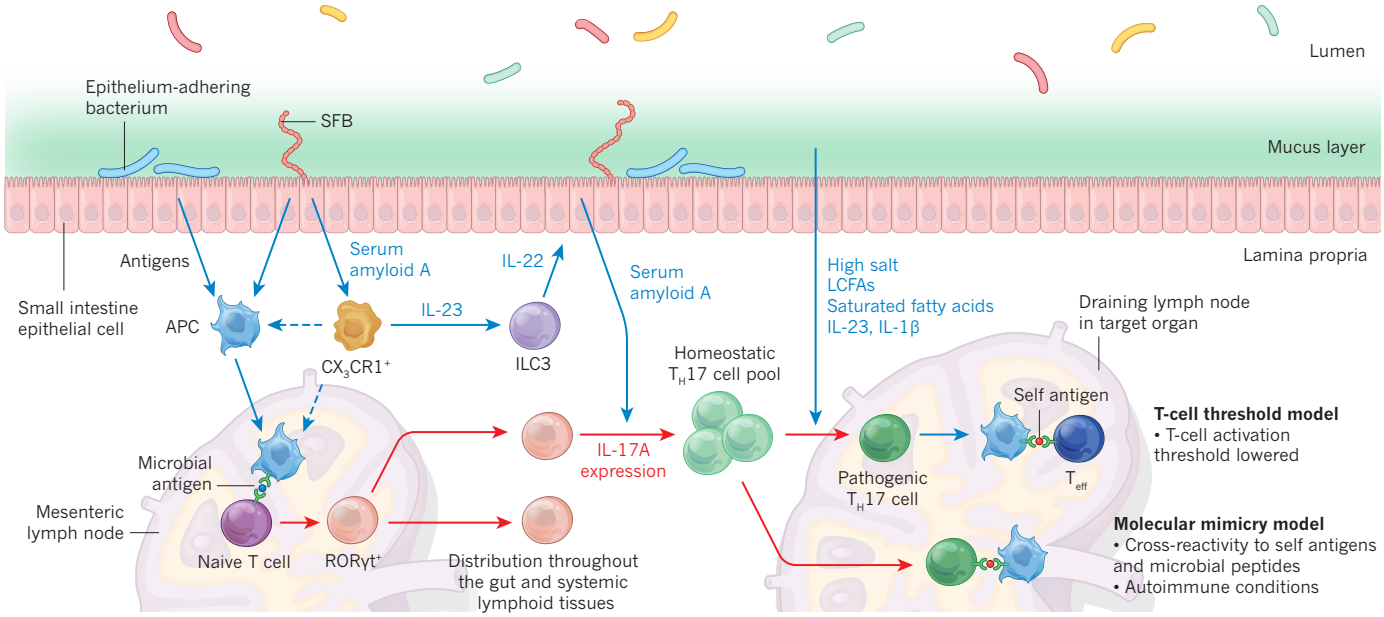


Figure 2 | Microbiota-mediated induction of T_H17 cells and autoimmunity. Epithelium-adhering bacteria initiate the differentiation of naive $CD4^+$ T cells into RORγt-expressing T cells (T_H17 polarized cells) (red) in the mesenteric lymph node through as-yet-undefined antigen-presenting cells (APCs). T_H17 polarized cells then accumulate and further differentiate into IL-17-expressing homeostatic T_H17 cells (green) in the lamina propria of the small intestine. These homeostatic T_H17 cells then stimulate epithelial cells to enhance the integrity of the intestinal mucosal barrier. The adhesion of segmented filamentous bacteria (SFB) elicits a unique program of gene expression in the epithelial cells, including the upregulation of serum amyloid A. Serum amyloid A from epithelial cells of the small intestine seems to function as a cytokine and it modulates CX_3CR1-expressing cells (that are derived from monocytes) to produce IL-23, which stimulates the production of IL-22 by ILC3s. As well as its effects on CX_3CR1-expressing cells, serum amyloid A can stimulate RORγt-expressing T cells directly to upregulate the expression of IL-17A. Dendritic cells that express the antigens CD11b and CD103 have also been implicated in the expansion and maintenance of T_H17 cells (not shown). T_H17 cells become pathogenic when they are stimulated with IL-23, IL-1β, higher concentrations of salt, long-chain fatty acids (LCFAs) and saturated fatty acids. Pathogenic T_H17 cells can migrate to the draining lymph nodes of target organs, where they contribute to autoimmune disease through cross-reactivity between peptides from microbes and self antigens (the molecular mimicry model). Alternatively, microbiota-specific T_H17 cells migrate to the lymph nodes and lower the threshold of activation of auto-reactive T cells such as T_{eff} cells (the T-cell threshold model).

migrate out of the intestine and into the spleen[64]. Self-reactive but gut-microbiota-activated T_H17 cells might contribute to other autoimmune disorders, including uveitis[65] and encephalomyelitis[66]. Such T-cell-mediated autoimmune conditions could be caused by cross-reactivity between microbial peptides and self antigens[67], a process known as molecular mimicry (Fig. 2). This model is consistent with the fact that the genes of the major histocompatibility complex (MHC) are the most important genetic susceptibility loci for many autoimmune disorders. Alternatively, microbiota-specific T_H17 cells might mediate some kind of bystander effect. This is because autoimmune disorders often affect more than one organ, and the genes that encode the signalling molecules that act downstream of TCRs are important determinants of genetic susceptibility to various autoimmune disorders in humans, including rheumatoid arthritis[68]. The T-cell threshold model proposes that gut-microbiota-activated T_H17 cells might migrate into the draining lymph nodes of the target organs and either lower the threshold of activation of autoreactive T cells or have their own activation threshold lowered. Indeed, T_H17 cells that are specific to SFB and that are primed in gut-draining lymph nodes can be found in other lymph nodes and in the spleen[59]. When produced aberrantly in some organs, molecules such as serum amyloid A might serve an adjuvant function and contribute to the heightened activity of such T cells (Fig. 2).

The potential for detrimental inflammation suggests that the responses of T cells and B cells to the gut microbiota must be tightly regulated. This is achieved through a number of mechanisms, including T-cell depletion and anergy. In this context, expression of MHC class II molecules by ILC3s has been found to restrain the expansion of T_H17 cells[69]. This could occur through the presentation of antigens that are derived from commensal bacteria to induce apoptosis of the antigen-specific T cells[69], although an antigen-presenting function for ILC3s is yet to be demonstrated. Beyond this context, however, one of the most crucial mechanisms for restraining inflammation in the gut is the induction of $CD4^+$ T_{reg} cells that express forkhead box protein P3 (Foxp3).

Induction of T_{reg} cells by the microbiota

T_{reg} cells that express both CD4 and Foxp3 can be found in every organ of the body, and they comprise a high proportion of the T cells of the lamina propria of the intestine[5,70–73]. Intestinal T_{reg} cells play an important part in maintaining immune tolerance to dietary antigens and the gut microbiota[74,74] as well as in suppressing tissue damage inflicted by immune responses against pathogenic bacteria such as *C. rodentium*[76] that are mediated by T_{eff} cells. The intestine contains both thymus-derived T_{reg} (tT_{reg}) cells and peripherally differentiated T_{reg} (pT_{reg}) cells; pT_{reg} cells are substantially enriched in the colon, mainly express RORγt and generally lack the zinc-finger protein Helios and the receptor neuropilin 1 (Nrp1) (refs 77–79) (Fig. 3). Because pT_{reg} cells disappear under germ-free conditions, they are probably induced by the microbiota[77–79]. Consistent with this, T_{reg} cells that express RORγt show the restricted TCR repertoire of cells that have proliferated in response to peripheral stimuli, but their TCR sequences overlap with those of $CD4^+$ T cells that lack Foxp3 (ref. 79). Experiments to track the fate of immature T cells that express a transgenic TCR cloned from colonic T_{reg} cells demonstrate that the expansion and differentiation of the transgenic T cells into T_{reg} cells occurs in the colon in the presence of cognate commensal bacteria and not in the thymus[80]. A considerable fraction of $ROR\gamma t^+$ T_{reg} cells express IL-10 (ref. 77), which is also produced by many other types of cell, including type 1 regulatory (Tr1) cells and myeloid cells, which have important roles in maintaining homeostasis in the intestines[81,82]. T_{reg}-cell-derived IL-10 is essential for suppression of the aberrant activation of myeloid cells, γδ T cells and T_H17 cells[83–85]. T_{reg} cells that express RORγt also express high levels of cytotoxic T-lymphocyte protein 4 (CTLA-4) (ref. 77) and are more effective than RORγt-negative T_{reg} cells in restraining immune pathogenesis

© 2016 Macmillan Publishers Limited. All rights reserved.

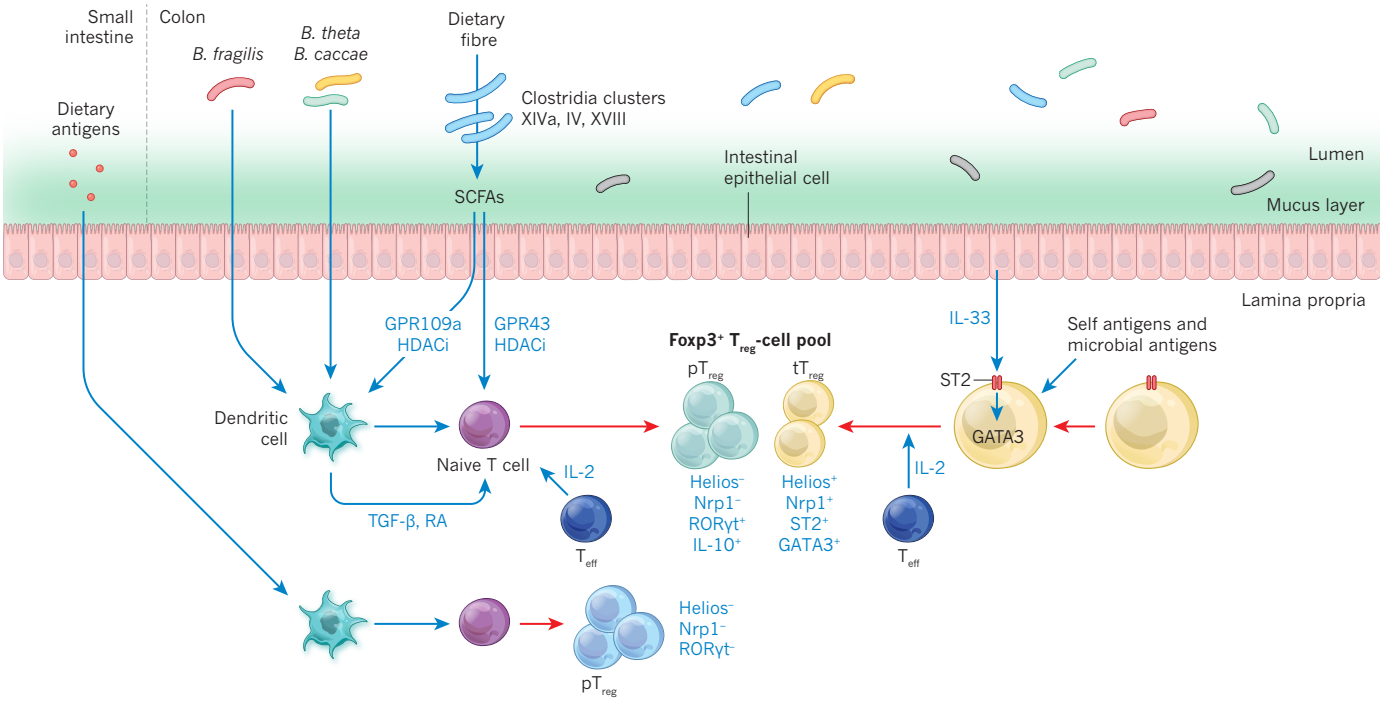


Figure 3 | Influence of the microbiota and diet on subsets of regulatory T cells in the intestine. Foxp3-expressing T_{reg} cells in the intestine can be subdivided into at least three subsets on the basis of their expression of RORγt, GATA3, Helios and Nrp1. T_{reg} cells that express RORγt but not Nrp1 are induced at peripheral sites by antigens derived from the microbiota. Known as pT_{reg} cells, they are the main producers of IL-10, which suppresses the aberrant activation of myeloid cells, γδ T cells and T_H17 cells. Dendritic cells produce mediators of pT_{reg}-cell differentiation, including TGF-β and retinoic acid (RA). Short-chain fatty acids (SCFAs), which are produced from dietary fibre by certain members of the microbiota, particularly species of Clostridia, also contribute to the induction of pT_{reg} cells. On binding to the G-protein-coupled receptor (GPR)109a on dendritic cells, short-chain fatty acids induce the expression of aldehyde dehydrogenase (ALDH), which metabolizes vitamin A into RA. SCFAs entering dendritic cells act as inhibitors of histone deacetylase (HDACi) to suppress the expression of pro-inflammatory cytokines. They also directly act on naive T cells through GPR43 or the upregulation of Foxp3 expression through HDAC inhibition. IL-2 derived from T_{eff} cells probably helps to stabilize the differentiation of T_{reg} cells. Several species of Bacteroides contribute to the induction of pT_{reg} cells that express RORγt but not Nrp1 through dendritic cells. A second pool of pT_{reg} cells expresses neither RORγt nor Nrp1; these T_{reg} cells are induced by, and maintain immune tolerance to, dietary antigens. It should be noted that induction of pT_{reg} cells through dietary antigens occurs largely in the small intestine, whereas the induction of pT_{reg} cells by the microbiota occurs largely in the colon. T_{reg} cells that express both GATA3 and Nrp1 are thought to be generated in the thymus and are known as tT_{reg} cells. GATA3$^+$ T_{reg} cells express ST2 (a component of the IL-33 receptor that is also known as IL1RL1). IL-33, which is probably released from the epithelial cells of the intestine at steady state, is markedly upregulated under conditions of inflammation. IL-33 acts with IL-2 (from T_{eff} cells) to induce the expression of GATA3 in T_{reg} cells.

in models of colitis[78,79]. Conditional inactivation of RORγt using the Cre–Lox recombination system in Foxp3$^+$ intestinal T cells in mice results in T_H2-cell-mediated inflammation[77] or in the expansion of T_H17 cells[78]. It should be noted that some intestinal T_H17 cells lose IL-17A expression in the presence of SFB and a fraction of these ex-T_H17 cells express Foxp3 (ref. 86). Foxp3$^+$ Cre–Lox mice in which RORγt has been inactivated might therefore reflect their RORγt deficiency in ex-T_H17 cells as well as microbiota-induced pT_{reg} cells.

The intestine also contains a subpopulation of T_{reg} cells that expresses the transcription factor GATA3 (ref. 87) (Fig. 3). These cells are distinct from RORγt$^+$ T_{reg} cells, and most express Nrp1 and Helios and are unaffected by the absence of the gut microbiota, which suggests that they mainly derive from tT_{reg} cells[78]. T_{reg} cells that express GATA3 co-express the IL-33 receptor ST2 (also known as IL1RL1) (ref. 88). IL-33, which is produced by the epithelial cells of the intestine at high levels under conditions of inflammation, works with IL-2 and the process of TCR engagement to induce the expression of GATA3 in T_{reg} cells. GATA3 upregulates the expression of Foxp3 and ST2 in a feed-forward process that promotes the proliferation and maintenance of T_{reg} cells[88]. T_{reg} cells that express Foxp3 but that lack RORγt and Nrp1 constitute one-third of the T_{reg}-cell population and are uniquely abundant in the lamina propria of the small intestine[89] (Fig. 3). This subpopulation is unaffected by the absence of the gut microbiota but disappears in germ-free mice that are fed an antigen-free diet[89]. Such cells therefore seem to be pT_{reg} cells that are induced by dietary antigens, and they constitute a subpopulation that can be distinguished from microbiota-induced, RORγt$^+$ pT_{reg} cells and from GATA3-expressing tT_{reg} cells. Mice that lack this subpopulation exhibit an increased susceptibility to food allergies[89]. Certain pT_{reg}-cell and tT_{reg}-cell subpopulations might have complementary and context-dependent functions, such as immune regulation at steady state in response to components of the microbiota (by RORγt$^+$ T_{reg} cells that lack Nrp1) and of the diet (by RORγt-negative T_{reg} cells that also lack Nrp1) and under conditions of inflammation that is triggered by self antigens (by GATA3$^+$ T_{reg} cells that express Nrp1).

The parts played by individual members or defined communities of the gut microbiota in the accumulation and functional maturation of T_{reg} cells of the intestine are starting to be illuminated. For example, strains that fall within clusters IV, XIVa and XVIII of Clostridia have a strong capacity for inducing the accumulation of T_{reg} cells in the colon[4,5] (Fig. 3). Oral administration to germ-free mice of a mixture of 46 strains of Clostridia that were derived from the faeces of conventional mice[90] leads to the strong induction of T_{reg} cells in the colon[5]. Similarly, a mixture of 17 strains of Clostridia that were isolated from a healthy person strongly induces T_{reg} cells in the colons of mice and rats[4]. This mixture preferentially enhances the accumulation of RORγt-expressing T_{reg} cells that lack Helios, rather than of GATA3-expressing T_{reg} cells[4,77]. Strains of Clostridia can also facilitate the expression of IL-10 and CTLA-4 by T_{reg} cells[4,5], and mice with an abundance of strains of Clostridia in their intestines exhibit resistance to experimental colitis[4,5]. In mouse models of graft-versus-host disease, the introduction of 17 strains of T_{reg}-inducing Clostridia reduces severity of the disease[91]. These Clostridia also stimulate ILC3s to produce IL-22, which helps to reinforce the epithelial barrier

© 2016 Macmillan Publishers Limited. All rights reserved.

and reduces the permeability of the intestine to dietary proteins[73]. Mice colonized by a microbiota that includes Clostridia therefore display a suppressed response to food allergens[73]. Clostridia-induced T_{reg} cells support the production of IgA in the intestine, which contributes to increased diversity of the microbiota and, in particular, of Clostridia[14].

One species of Clostridia, *Faecalibacterium prausnitzii*, is underrepresented in people with inflammatory bowel disease[92] and it promotes the accumulation of IL-10-expressing T cells that are positive for both CD4 and CD8αα in the colon[93]. A population of lymphocytes from the intestinal epithelium that is positive for both such antigens could have a similar immune regulatory role in the small intestine of the mouse. These microbiota-dependent T cells differentiate in the periphery on loss of the expression of the CD4-lineage transcription factor ThPOK and upregulation of the CD8-lineage transcription factor Runx3[94,95]. How these cells function in preventing the differentiation of inflammatory cells in the small intestine is yet to be determined.

A small consortium of microbes known as altered Schaedler flora, which contains strains of Clostridia, is also capable of increasing the number of T_{reg} cells in the lamina propria of the mouse colon[71]. The precise mechanism through which Clostridia stimulate the induction of T_{reg} cells in the colon remains to be elucidated. One possible mechanism is the cooperative production of short-chain fatty acids through fermentation of dietary fibre[4,39,96] (Fig. 3). For example, the collective genomes of the 17 strains of T_{reg}-cell-inducing Clostridia contain numerous genes that are predicted to be involved in the biosynthesis of short-chain fatty acids[96]. Short-chain fatty acids suppress the expression of pro-inflammatory cytokines in dendritic cells through the inhibition of histone deacetylases (HDACs)[40] and through the activation of the G-protein-coupled receptor (GPR)109a (also known as HCAR2) (ref. 97). They can also stimulate the proliferation of T_{reg} cells directly by activating GPR43 (FFAR2) (ref. 38) and the differentiation of naive $CD4^+$ T cells into pT_{reg} cells through HDAC inhibition, which results in histone H3 acetylation at the conserved non-coding sequence (CNS)1 element of the gene *Foxp3* (ref. 39). *In vitro* stimulation of T_{reg} cells with short-chain fatty acids upregulates the expression of GPR15 (ref. 38), which promotes the recruitment of T_{reg} cells to the colon[76], although this has not been demonstrated *in vivo*.

Programs for the induction of T_{reg} cells can also be activated by non-Clostridia members of the microbiota. *Lactobacillus reuteri* and *L. murinus* have been shown to increase the proportion of T_{reg} cells in mice[98–100]. Infection with *Helicobacter hepaticus* induces IL-10-producing T_{reg} cells that inhibit the development of colitis in an *H. hepaticus* antigen-specific manner[101]. *Bacteroides fragilis* boosts the production of IL-10 by T_{reg} cells of the colon, and this activity is mediated by polysaccharide A[70] from the bacterium's capsule. Outer-membrane vesicles containing polysaccharide A that are released by *B. fragilis* might also be taken up by dendritic cells of the intestine to stimulate their production of IL-10 through TLR-2 signalling[102]. The IL-10 from these dendritic cells might then induce T_{reg} cells to also produce IL-10. Several other species of *Bacteroides*, including *B. caccae* and *B. thetaiotaomicron*, also induce the accumulation of $Foxp3^+$ cells, particularly RORγt-expressing pT_{reg} cells in the colon[78,103]. Collectively, there is considerable overlap between the responses of T_{reg} cells to Clostridia, *Lactobacillus* and *Bacteroides*, which indicates that different pathways for the regulation of T_{reg} cells converge in the intestine. The induction and maintenance of T_{reg} cells might be a common and crucial mechanism for maintaining the homeostatic and beneficial relationship between the microbiota and the host.

It has been suggested that tolerogenic dendritic cells that carry the CD103 antigen contribute to the induction of T_{reg} cells[104,105]. CSF2 that is produced in response to the microbiota by ILC3s might also act on dendritic cells of the colon to promote the expansion of T_{reg} cells[106] (Box 1) (Fig. 3). The ablation of MHC class II expression in conventional dendritic cells, which include $CD103^+$ dendritic cells, results in reduced induction of pT_{reg} cells and in spontaneous inflammation[107]. The TCR repertoires of pT_{reg} cells and tT_{reg} cells differ substantially. In one study, at least half of the TCRs that were cloned from T_{reg} cells of the colon and expressed in a reporter hybridoma cell line responded to autoclaved contents of the intestines of mice, and two TCR clones were stimulated by strains of *Parabacteroides distasonis* or by an uncharacterized species of Clostridia[80]. Consistent with this, at least some of the T_{reg} cells that were induced by the mixture of 17 strains of human-derived Clostridia reacted to Clostridia antigens[4]. Whether there is a role for antigen specificity in T_{reg}-cell-mediated tolerance at mucosal surfaces, however, is an important question that still needs to be answered.

Cells from the adaptive immune system that are primed at microbiota-sensing mucosa can take up residence in and protect other mucosal surfaces. For example, intranasal vaccination is particularly effective at eliciting protective memory-T-cell responses against *Chlamydia trachomatis* in the female reproductive tract[108]. However, when ultraviolet-inactivated *C. trachomatis* is delivered intramucosally, antigens accumulate preferentially in CD103-expressing dendritic cells that lack CD11b and induce antigen-specific T_{reg} cells, and no protective immunity is elicited[108]. By contrast, immunization with ultraviolet-inactivated *C. trachomatis* conjugated to adjuvant nanoparticles that target CD103-lacking dendritic cells that express CD11b provides effective, antigen-specific memory responses by mucosa-resident T_H1 cells[108]. The immune responses that mucosal bacteria elicit therefore differ according to the route and context of antigen delivery. Elucidation of the mechanisms by which commensal microbes deliver antigens for presentation in an immunogenic versus a tolerogenic context might enable the development of effective mucosal vaccines.

Implications for disease and therapeutics

Members of the gut microbiota have distinct effects on homeostasis of the host's adaptive immune system. Differences in the composition of the community therefore contribute to variability in immune responses and susceptibility to infection, autoimmune disorders, allergy and other immunological conditions. Understanding the development of the mucosal immune system and its dysregulation in relation to normal and dysbiotic microbiotas is important for the development of drugs, probiotic supplements, vaccines and cancer immunotherapies.

A crucial window of time

The microbiota is established in early life. Indeed, an absence of microbiota during this period of development leads to increases in the number of invariant natural killer T (iNKT) cells and in susceptibility to colitis and asthma[109] in animal models. Early exposure to the gut microbiota suppresses the abundance of iNKT cells in the gut and lung, partly through the epigenetic suppression of the gene that encodes the chemokine CXCL16 (ref. 109). Colonization with commensal bacteria during the neonatal period also results in the recruitment of T_{reg} cells to mucosal sites and the establishment of long-lasting tolerance to the microbes[110]. Treatment with antibiotics results in an increase in susceptibility to asthma in perinatal, but not adult, mice through a decrease in the accumulation of T_{reg} cells in the colon and an enhanced IgE response[111]. In the absence of colonization by a microbiota at an early age, B cells preferentially undergo isotype switching to IgE, rather than IgA[112,113]. The elevated concentration of IgE in the blood serum of germ-free mice is accompanied by an increase in circulating basophils and exaggerated basophil-mediated T_H2-cell responses and allergic inflammation[112]. The induction of IgE is not suppressed by colonization with a microbiota later in life or by early colonization with a microbiota of limited complexity[113].

A cohort of children who had a high risk of developing atopy and asthma were found to have microbiotal dysbiosis that is characterized by a reduction in four specific genera of bacteria: *Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia* — collectively known as FLVR[114]. Colonization of mice with FLVR mitigated airway inflammation in a model of allergic asthma, which raises the prospect that atopy or asthma could be averted by early therapy to correct dysbiosis[114]. There is a crucial window of time in early life, therefore, during which exposure to diverse microbiota is extremely important for the suppression of iNKT cells and IgE-expressing cells, the induction and expansion of T_{reg} cells

© 2016 Macmillan Publishers Limited. All rights reserved.

and the establishment of systemic tolerance to a large spectrum of environmental antigens.

Dysbiosis and aberrant adaptive immunity

Microbiotal dysbiosis can be caused by genetic predisposition, infections and changes in diet and nutritional status, as well as the use of antibiotics, agents that suppress gastric acid and anticancer drugs. Although there is convincing evidence to suggest that dysbiosis causes or promotes disease, the underlying mechanisms are not fully understood. Several reports describe the association of particular species of bacteria with autoimmune and inflammatory conditions[115,116]. In a mouse model, the administration of a diet rich in milk fat induces a bloom of taurocholic-acid-consuming *Bilophila wadsworthia*, which enhances the response of T_H1 cells and accelerates the onset of colitis[115]. Adherent-invasive *E. coli* (AIEC) is frequently observed in people with inflammatory bowel disease and can induce an active response by T_H17 cells in mice[116]. Mutations in the gene *NOD2*, found in subsets of people with inflammatory bowel disease, are associated with shifts in the composition of the gut microbiota[117]. *Nod2* deficiency in mice results in the expansion of the commensal bacterium *Bacteroides vulgatus*, which is accompanied by an excessive IFN-γ response from intraepithelial lymphocytes[118]. Colonization of the intestinal mucosa by bacteria from the mouth, such as Veillonellaceae and Fusobacteriaceae, is one of the earliest events in children with new-onset Crohn's disease[119]. Similarly, there is an increased prevalence of *Prevotella copri* in the faecal microbiota of people with new-onset rheumatoid arthritis[120]. However, the ability of these bacteria to trigger disease is yet to be established.

As well as the activation of T_{eff} cells in response to potentially pathogenic bacteria, compromised barrier function of the epithelium and dysregulated responses to the commensal microbiota are important features of chronic inflammatory diseases that are associated with dysbiosis. For instance, infection with HIV leads to chronic dysbiosis with a reduction in Clostridia and Bacteroidia and an enrichment of taxa that produce enzymes for tryptophan catabolism, and is accompanied by heightened permeability of the mucosa, elevated levels of T-cell activation and diminished numbers of IL-17-secreting mucosal T cells[121]. These events might contribute collectively to the chronic inflammation that is observed in individuals who are infected with HIV. In a mouse model, infection with the protozoa *Toxoplasma gondii* and subsequent disruption of the epithelial barrier induces memory T_H1 cells specific for commensal Clostridia that normally induce T_{reg} cells and IgA-secreting B cells[122,123]. Similarly, responses to flagellin antigens (known as CBir) that are expressed by commensal species from the Clostridia cluster XIVa have been detected in people with Crohn's disease[124]. Importantly, the transfer of CD4-expressing T cells that are specific for CBir into immunodeficient mice that have been colonized with commensal Clostridia causes severe colitis[124]. Disruption of the epithelial barrier owing to the complex interplay between a dysbiotic microbiota and pathogenic bacteria might therefore lead to dysregulated immune responses to commensal microbes, chronic inflammation and the stabilization of a pro-inflammatory community of microbes.

Cancer immunotherapy

The importance of the composition of the microbiota in how tumour-carrying hosts respond to chemotherapy or checkpoint blockade immunotherapy has been highlighted in several studies. Reductions in the growth of sarcomas in mice following treatment with the chemotherapeutic drug cyclophosphamide can be compromised after exposure to antibiotics, and this has been attributed to the loss of anti-tumour T_H17-cell-inducing commensal bacteria, the growth of which is favoured by the chemotherapy[125]. However, it is unknown whether the beneficial anti-tumour properties of microbiota-dependent T_H17 cells are broadly applicable. Similarly, antibiotics compromise the anti-tumour response that follows CTLA-4 blockade in mice[126]. In this case, anti-CTLA-4 immunotherapy favours the dominance of species of *Bacteroides*, such as *B. fragilis* and *B. thetaiotaomicron*, in both mice and people. These bacteria are of benefit because they enhance the efficacy of the CTLA-4 blockade, possibly through an anti-tumour response mediated by T_H1 cells[126]. In another mouse model, colonization of the gut with Bifidobacteria has been found to contribute to the control of implanted syngeneic tumours by CD8-expressing T cells following anti-PD-L1 cancer immunotherapy[3]. The mechanism for the improved anti-tumour response might involve activation of the functions of antigen-presenting cells, followed by improved infiltration of tumours by cytotoxic T cells, although it remains to be determined whether microbiota-regulated $CD4^+$ T cells also have a role in restraining the growth of tumours.

Outlook

Studies of how the mutualistic relationship between cells of the adaptive immune system and members of the microbiota affect health and disease are in their infancy. Most efforts have strived to establish reductionist approaches that can be exploited to elucidate cellular and molecular mechanisms. From a translational perspective, models of humanized microbiota in germ-free mice and pigs have been established[127]. It is possible that these efforts will permit the design of bacterial consortia and metabolic products that durably activate or suppress specific programs of adaptive immunity, which will result in the development of improved vaccines and therapeutic drugs for disorders that involve the immune system — including infections, autoimmunity, allergies and cancer. It should be noted, however, that the interactions between the microbiota and the host are influenced to a large extent by host genetics, cooperation and competition between pathogenic and commensal microbes and multiple environmental variables, including diet, circadian factors and the climate. The 'one microbe, one response' approach will probably need to be supplanted by more integrative systems analyses that require the development of advanced technologies and computational tools. Improved characterization of metabolites or other microbial effectors, coupled with computational pathway analyses, might enable the design of synthetic organisms or postbiotic products that can shape immune responses. Elucidation of the role of viruses and phages might provide further approaches for targeting components of the microbiota or host cells for therapeutic purposes. The role of the microbiota in shaping adaptive immunity should therefore become an increasingly fertile area for basic and translational investigation. ■

Received 21 February; accepted 25 April 2016.

1. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139,** 485–498 (2009).
   Together with ref. 50, this study shows that a subset of the microbiota specifically affects the accumulation of T_H17 cells in the intestine.
2. Wu, H.-J. *et al.* Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* **32,** 815–827 (2010).
3. Sivan, A. *et al.* Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* **350,** 1084–1089 (2015).
   Together with refs 125 and 126, this study shows that a subset of the microbiota can have an effect on the efficacy of cancer therapy.
4. Atarashi, K. *et al.* T_{reg} induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500,** 232–236 (2013).
   This study and ref. 5 show that a rationally selected consortium of bacteria can specifically induce T_{reg} cells in the intestine that function in systemic immune regulation.
5. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331,** 337–341 (2011).
6. Kau, A. L. *et al.* Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Sci. Transl. Med.* **7,** 276ra24 (2015).
   Together with refs 7 and 8, this study shows that IgA-SEQ is a powerful technique for identifying taxa that provide a strong stimulus to the host's immune system.
7. Palm, N. W. *et al.* Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* **158,** 1000–1010 (2014).
8. Bunker, J. J. *et al.* Innate and adaptive humoral responses coat distinct commensal bacteria with immunoglobulin A. *Immunity* **43,** 541–553 (2015).
9. Beura, L. K. *et al.* Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* **532,** 512–516 (2016).
10. Roche, A. M., Richard, A. L., Rahkola, J. T., Janoff, E. N. & Weiser, J. N. Antibody blocks acquisition of bacterial colonization through agglutination. *Mucosal Immunol.* **8,** 176–185 (2015).
11. Pabst, O. New concepts in the generation and functions of IgA. *Nature Rev. Immunol.* **12,** 821–832 (2012).

© 2016 Macmillan Publishers Limited. All rights reserved.

12. Peterson, D. A., McNulty, N. P., Guruge, J. L. & Gordon, J. I. IgA response to symbiotic bacteria as a mediator of gut homeostasis. *Cell Host Microbe* **2,** 328–339 (2007).
13. Cullender, T. C. *et al.* Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. *Cell Host Microbe* **14,** 571–581 (2013).
14. Kawamoto, S. *et al.* Foxp3$^+$ T cells regulate immunoglobulin A selection and facilitate diversification of bacterial species responsible for immune homeostasis. *Immunity* **41,** 152–165 (2014).
15. Friman, V., Nowrouzian, F., Adlerberth, I. & Wold, A. E. Increased frequency of intestinal *Escherichia coli* carrying genes for S fimbriae and haemolysin in IgA-deficient individuals. *Microb. Pathog.* **32,** 35–42 (2002).
16. Wei, M. *et al.* Mice carrying a knock-in mutation of *Aicda* resulting in a defect in somatic hypermutation have impaired gut homeostasis and compromised mucosal defense. *Nature Immunol.* **12,** 264–270 (2011).
17. Moon, C. *et al.* Vertically transmitted faecal IgA levels determine extra-chromosomal phenotypic variation. *Nature* **521,** 90–93 (2015).
18. Kubinak, J. L. *et al.* MyD88 signaling in T cells directs IgA-mediated control of the microbiota to promote health. *Cell Host Microbe* **17,** 153–163 (2015).
19. Hirota, K. *et al.* Plasticity of T_H17 cells in Peyer's patches is responsible for the induction of T cell-dependent IgA responses. *Nature Immunol.* 14, 372–379 (2013).
20. Hapfelmeier, S. *et al.* Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science* **328,** 1705–1709 (2010).
21. Lindner, C. *et al.* Diversification of memory B cells drives the continuous adaptation of secretory antibodies to gut microbiota. *Nature Immunol.* **16,** 880–888 (2015).
22. Ivanov, I. I. *et al.* The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17$^+$ T helper cells. *Cell* **126,** 1121–1133 (2006).
23. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4,** 337–349 (2008).
24. Atarashi, K. *et al.* ATP drives lamina propria T_H17 cell differentiation. *Nature* **455,** 808–812 (2008).
25. Weaver, C. T., Elson, C. O., Fouser, L. A. & Kolls, J. K. The Th17 pathway and inflammatory diseases of the intestines, lungs, and skin. *Annu. Rev. Pathol.* **8,** 477–512 (2013).
26. Puel, A. *et al.* Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science* **332,** 65–68 (2011).
27. Okada, S. *et al.* Impairment of immunity to *Candida* and *Mycobacterium* in humans with bi-allelic *RORC* mutations. *Science* **349,** 606–613 (2015).
28. Ishigame, H. *et al.* Differential roles of interleukin-17A and -17F in host defense against mucoepithelial bacterial infection and allergic responses. *Immunity* **30,** 108–119 (2009).
29. McGeachy, M. J. *et al.* The interleukin 23 receptor is essential for the terminal differentiation of interleukin 17-producing effector T helper cells *in vivo*. *Nature Immunol.* **10,** 314–324 (2009).
30. Coccia, M. *et al.* IL-1β mediates chronic intestinal inflammation by promoting the accumulation of IL-17A secreting innate lymphoid cells and CD4$^+$ Th17 cells. *J. Exp. Med.* **209,** 1595–1609 (2012).
31. Hirota, K. *et al.* Fate mapping of IL-17-producing T cells in inflammatory responses. *Nature Immunol.* **12,** 255–263 (2011).
32. El-Behi, M. *et al.* The encephalitogenicity of T_H17 cells is dependent on IL-1- and IL-23-induced production of the cytokine GM-CSF. *Nature Immunol.* **12,** 568–575 (2011).
33. Harbour, S. N., Maynard, C. L., Zindl, C. L., Schoeb, T. R. & Weaver, C. T. Th17 cells give rise to Th1 cells that are required for the pathogenesis of colitis. *Proc. Natl Acad. Sci. USA* **112,** 7061–7066 (2015).
34. Ahern, P. P. *et al.* Interleukin-23 drives intestinal inflammation through direct activity on T cells. *Immunity* **33,** 279–288 (2010).
35. Jain, R. *et al.* Interleukin-23-induced transcription factor Blimp-1 promotes pathogenicity of T helper 17 cells. *Immunity* **44,** 131–142 (2016).
36. Wu, C. *et al.* Induction of pathogenic T_H17 cells by inducible salt-sensing kinase SGK1. *Nature* **496,** 513–517 (2013).
37. Kleinewietfeld, M. *et al.* Sodium chloride drives autoimmune disease by the induction of pathogenic T_H17 cells. *Nature* **496,** 518–522 (2013).
38. Smith, P. M. *et al.* The microbial metabolites, short-chain fatty acids, regulate colonic T_{reg} cell homeostasis. *Science* **341,** 569–573 (2013).
Together with refs 39–41, this study identified short-chain fatty acids as strong inducers of T_{reg} cells in the colon.
39. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504,** 446–450 (2013).
40. Arpaia, N. *et al.* Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* **504,** 451–455 (2013).
41. Haghikia, A. *et al.* Dietary fatty acids directly impact central nervous system autoimmunity via the small intestine. *Immunity* **43,** 817–829 (2015).
42. Berod, L. *et al. De novo* fatty acid synthesis controls the fate between regulatory T and T helper 17 cells. *Nature Med.* **20,** 1327–1333 (2014).
43. Santori, F. R. *et al.* Identification of natural RORγ ligands that regulate the development of lymphoid cells. *Cell Metab.* **21,** 286–297 (2015).
44. Wang, C. *et al.* CD5L/AIM regulates lipid biosynthesis and restrains Th17 cell pathogenicity. *Cell* **163,** 1413–1427 (2015).
45. Naik, S. *et al.* Compartmentalized control of skin immunity by resident commensals. *Science* **337,** 1115–1119 (2012).
46. Umesaki, Y., Setoyama, H., Matsumoto, S., Imaoka, A. & Itoh, K. Differential roles of segmented filamentous bacteria and clostridia in development of the intestinal immune system. *Infect. Immun.* **67,** 3504–3511 (1999).
47. Lécuyer, E. *et al.* Segmented filamentous bacterium uses secondary and tertiary lymphoid tissues to induce gut IgA and specific T helper 17 cell responses. *Immunity* **40,** 608–620 (2014).
48. Goto, Y. *et al.* Innate lymphoid cells regulate intestinal epithelial cell glycosylation. *Science* **345,** 1254009 (2014).
49. Prakash, T. *et al.* Complete genome sequences of rat and mouse segmented filamentous bacteria, a potent inducer of Th17 cell differentiation. *Cell Host Microbe* **10,** 273–284 (2011).
50. Atarashi, K. *et al.* Th17 cell induction by adhesion of microbes to intestinal epithelial cells. *Cell* **163,** 367–380 (2015).
Together with ref. 51, this study shows that the response of intestinal T_H17 cells is directed towards commensal and pathogenic bacteria that activate epithelial cells.
51. Sano, T. *et al.* An IL-23R/IL-22 circuit regulates epithelial serum amyloid A to promote local effector Th17 responses. *Cell* **163,** 381–393 (2015); erratum **164,** 324 (2016).
52. Schnupf, P. *et al.* Growth and host interaction of mouse segmented filamentous bacteria *in vitro*. *Nature* **520,** 99–103 (2015).
53. Panea, C. *et al.* Intestinal monocyte-derived macrophages control commensal-specific Th17 responses. *Cell Rep.* **12,** 1314–1324 (2015).
54. Lewis, K. L. *et al.* Notch2 receptor signaling controls functional differentiation of dendritic cells in the spleen and intestine. *Immunity* **35,** 780–791 (2011).
55. Persson, E. K. *et al.* IRF4 transcription-factor-dependent CD103$^+$CD11b$^+$ dendritic cells drive mucosal T helper 17 cell differentiation. *Immunity* **38,** 958–969 (2013).
56. Schlitzer, A. *et al.* IRF4 transcription factor-dependent CD11b$^+$ dendritic cells in human and mouse control mucosal IL-17 cytokine responses. *Immunity* **38,** 970–983 (2013).
57. Derebe, M. G. *et a*l. Serum amyloid A is a retinol binding protein that transports retinol during bacterial infection. *eLife* **3,** e03206 (2014).
58. Sczesnak, A. *et al.* The genome of Th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* **10,** 260–272 (2011).
59. Yang, Y. *et al.* Focused specificity of intestinal T_H17 cells towards commensal bacterial antigens. *Nature* **510,** 152–156 (2014).
This study and ref. 122 show that different constituents of the microbiota guide distinct pathways of T-cell differentiation that is specific for the antigens of commensal bacteria.
60. Block, K. E., Zheng, Z., Dent, A. L., Kee, B. L. & Huang, H. Gut microbiota regulates K/BxN autoimmune arthritis through follicular helper T but not Th17 cells. *J. Immunol.* **196,** 1550–1557 (2016).
61. Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4615–4622 (2011).
62. Kriegel, M. A. *et al.* Naturally transmitted segmented filamentous bacteria segregate with diabetes protection in nonobese diabetic mice. *Proc. Natl Acad. Sci. USA* **108,** 11548–11553 (2011).
63. Fransen, F. *et al.* BALB/c and C57BL/6 mice differ in polyreactive IgA abundance, which impacts the generation of antigen-specific IgA and microbiota diversity. *Immunity* **43,** 527–540 (2015).
64. Morton, A. M. *et al.* Endoscopic photoconversion reveals unexpectedly broad leukocyte trafficking to and from the gut. *Proc. Natl Acad. Sci. USA* **111,** 6696–6701 (2014).
65. Horai, R. *et al.* Microbiota-dependent activation of an autoreactive T cell receptor provokes autoimmunity in an immunologically privileged site. *Immunity* **43,** 343–353 (2015).
66. Berer, K. *et al.* Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* **479,** 538–541 (2011).
67. Harkiolaki, M. *et al.* T cell-mediated autoimmune disease due to low-affinity crossreactivity to common microbial peptides. *Immunity* **30,** 348–357 (2009).
68. Sakaguchi, N. *et al.* Altered thymic T-cell selection due to a mutation of the *ZAP-70* gene causes autoimmune arthritis in mice. *Nature* **426,** 454–460 (2003).
69. Hepworth, M. R. *et al.* Group 3 innate lymphoid cells mediate intestinal selection of commensal bacteria-specific CD4$^+$ T cells. *Science* **348,** 1031–1035 (2015).
70. Round, J. L. & Mazmanian, S. K. Inducible Foxp3$^+$ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **107,** 12204–12209 (2010).
71. Geuking, M. B. *et al.* Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity* **34,** 794–806 (2011).
72. Weiss, J. M. *et al.* Neuropilin 1 is expressed on thymus-derived natural regulatory T cells, but not mucosa-generated induced Foxp3$^+$ T reg cells. *J. Exp. Med.* **209,** 1723–1742 (2012).
73. Stefka, A. T. *et al.* Commensal bacteria protect against food allergen sensitization. *Proc. Natl Acad. Sci. USA* **111,** 13145–13150 (2014).
74. Bilate, A. M. & Lafaille, J. J. Induced CD4$^+$Foxp3$^+$ regulatory T cells in immune tolerance. *Annu. Rev. Immunol.* **30,** 733–758 (2012).
75. Josefowicz, S. Z., Lu, L. F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30,** 531–564 (2012).
76. Kim, S. V. *et al.* GPR15-mediated homing controls immune homeostasis in the large intestine mucosa. *Science* **340,** 1456–1459 (2013).
77. Ohnmacht, C. *et al.* The microbiota regulates type 2 immunity through RORγt$^+$ T cells. *Science* **349,** 989–993 (2015).
Together with refs 78 and 79, this study shows that a subset of T_{reg} cells in the intestine express RORγt and that their development is affected by the microbiota.
78. Sefik, E. *et al.* Individual intestinal symbionts induce a distinct population of

© 2016 Macmillan Publishers Limited. All rights reserved.

RORγ$^+$ regulatory T cells. *Science* **349,** 993–997 (2015).
79. Yang, B. H. *et al.* Foxp3 T cells expressing RORγt represent a stable regulatory T-cell effector lineage with enhanced suppressive capacity during intestinal inflammation. *Mucosal Immunol.* **9,** 444–457 (2016).
80. Lathrop, S. K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478,** 250–254 (2011).
81. Roers, A. *et al.* T cell-specific inactivation of the interleukin 10 gene in mice results in enhanced T cell responses but normal innate responses to lipopolysaccharide or skin irritation. *J. Exp. Med.* **200,** 1289–1297 (2004).
82. Krause, P. *et al.* IL-10-producing intestinal macrophages prevent excessive antibacterial innate immunity by limiting IL-23 synthesis. *Nature Commun.* **6,** 7055 (2015).
83. Rubtsov, Y. P. *et al.* Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28,** 546–558 (2008).
84. Huber, S. *et al.* Th17 cells express interleukin-10 receptor and are controlled by Foxp3$^-$ and Foxp3$^+$ regulatory CD4$^+$ T cells in an interleukin-10-dependent manner. *Immunity* **34,** 554–565 (2011).
85. Park, S. G. *et al.* T regulatory cells maintain intestinal homeostasis by suppressing γδ T cells. *Immunity* **33,** 791–803 (2010).
86. Gagliani, N. *et al.* T_H17 cells transdifferentiate into regulatory T cells during resolution of inflammation. *Nature* **523,** 221–225 (2015).
87. Wohlfert, E. A. *et al.* GATA3 controls Foxp3$^+$ regulatory T cell fate during inflammation in mice. *J. Clin. Invest.* **121,** 4503–4515 (2011).
88. Schiering, C. *et al.* The alarmin IL-33 promotes regulatory T-cell function in the intestine. *Nature* **513,** 564–568 (2014).
89. Kim, K. S. *et al.* Dietary antigens limit mucosal immunity by inducing regulatory T cells in the small intestine. *Science* **351,** 858–863 (2016).
90. Itoh, K. & Mitsuoka, T. Characterization of Clostridia isolated from faeces of limited flora mice and their effect on caecal size when associated with germ-free mice. *Lab. Anim.* **19,** 111–118 (1985).
91. Mathewson, N. D. *et al.* Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease. *Nature Immunol.* **17,** 505–513 (2016).
92. Sokol, H. *et al. Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105,** 16731–16736 (2008).
93. Sarrabayrouse, G. *et al.* CD4CD8αα lymphocytes, a novel human regulatory T cell subset induced by colonic bacteria and deficient in patients with inflammatory bowel disease. *PLoS Biol.* **12,** e1001833 (2014).
94. Reis, B. S., Rogoz, A., Costa-Pinto, F. A., Taniuchi, I. & Mucida, D. Mutual expression of the transcription factors Runx3 and ThPOK regulates intestinal CD4$^+$ T cell immunity. *Nature Immunol.* **14,** 271–280 (2013).
95. Mucida, D. *et al.* Transcriptional reprogramming of mature CD4$^+$ helper T cells generates distinct MHC class II-restricted cytotoxic T lymphocytes. *Nature Immunol.* **14,** 281–289 (2013).
96. Narushima, S. *et al.* Characterization of the 17 strains of regulatory T cell-inducing human-derived Clostridia. *Gut Microbes* **5,** 333–339 (2014).
97. Singh, N. *et al.* Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40,** 128–139 (2014).
98. Di Giacinto, C., Marinaro, M., Sanchez, M., Strober, W. & Boirivant, M. Probiotics ameliorate recurrent Th1-mediated murine colitis by inducing IL-10 and IL-10-dependent TGF-β-bearing regulatory cells. *J. Immunol.* **174,** 3237–3246 (2005).
99. Karimi, K., Inman, M. D., Bienenstock, J. & Forsythe, P. *Lactobacillus reuteri*-induced regulatory T cells protect against an allergic airway response in mice. *Am. J. Respir. Crit. Care Med.* **179,** 186–193 (2009).
100. Tang, C. *et al.* Inhibition of Dectin-1 signaling ameliorates colitis by inducing *Lactobacillus*-mediated regulatory T cell expansion in the intestine. *Cell Host Microbe* **18,** 183–197 (2015).
101. Kullberg, M. C. *et al.* Bacteria-triggered CD4$^+$ T regulatory cells suppress *Helicobacter hepaticus*-induced colitis. *J. Exp. Med.* **196,** 505–515 (2002).
102. Shen, Y. *et al.* Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. *Cell Host Microbe* **12,** 509–520 (2012).
103. Faith, J. J., Ahern, P. P., Ridaura, V. K., Cheng, J. & Gordon, J. I. Identifying gut microbe–host phenotype relationships using combinatorial communities in gnotobiotic mice. *Sci. Transl. Med.* **6,** 220ra11 (2014).
104. Coombes, J. L. *et al.* A functionally specialized population of mucosal CD103$^+$ DCs induces Foxp3$^+$ regulatory T cells via a TGF-β and retinoic acid-dependent mechanism. *J. Exp. Med.* **204,** 1757–1764 (2007).
105. Sun, C. M. *et al.* Small intestine lamina propria dendritic cells promote *de novo* generation of Foxp3 T reg cells via retinoic acid. *J. Exp. Med.* **204,** 1775–1785 (2007).
106. Mortha, A. *et al.* Microbiota-dependent crosstalk between macrophages and ILC3 promotes intestinal homeostasis. *Science* **343,** 1249288 (2014).
107. Loschko, J. *et al.* Absence of MHC class II on cDCs results in microbial-dependent intestinal inflammation. *J. Exp. Med.* **213,** 517–534 (2016).
108. Stary, G. *et al.* A mucosal vaccine against *Chlamydia trachomatis* generates two waves of protective memory T cells. *Science* **348,** aaa8205 (2015).
109. Olszak, T. *et al.* Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336,** 489–493 (2012).
110. Scharschmidt, T. C. *et al.* A wave of regulatory T cells into neonatal skin mediates tolerance to commensal microbes. *Immunity* **43,** 1011–1021 (2015).
111. Russell, S. L. *et al.* Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep.* **13,** 440–447 (2012).
112. Hill, D. A. *et al.* Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. *Nature Med.* **18,** 538–546 (2012).
113. Cahenzli, J., Koller, Y., Wyss, M., Geuking, M. B. & McCoy, K. D. Intestinal microbial diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe* **14,** 559–570 (2013).
114. Arrieta, M. C. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med.* **7,** 307ra152 (2015).
115. Devkota, S. *et al.* Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10*$^{-/-}$ mice. *Nature* **487,** 104–108 (2012).
116. Small, C. L., Reid-Yu, S. A., McPhee, J. B. & Coombes, B. K. Persistent infection with Crohn's disease-associated adherent-invasive *Escherichia coli* leads to chronic inflammation and intestinal fibrosis. *Nature Commun.* **4,** 1957 (2013).
117. Frank, D. N. *et al.* Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm. Bowel Dis.* **17,** 179–184 (2011).
118. Ramanan, D., Tang, M. S., Bowcutt, R., Loke, P. & Cadwell, K. Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity* **41,** 311–324 (2014).
119. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15,** 382–392 (2014).
120. Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2,** e01202 (2013).
121. Vujkovic-Cvijin, I. *et al.* Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism. *Sci. Transl. Med.* **5,** 193ra91 (2013).
122. Hand, T. W. *et al.* Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science* **337,** 1553–1556 (2012).
123. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T. R. & Elson, C. O. A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **106,** 19256–19261 (2009).
124. Lodes, M. J. *et al.* Bacterial flagellin is a dominant antigen in Crohn disease. *J. Clin. Invest.* **113,** 1296–1306 (2004).
125. Viaud, S. *et al.* The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science* **342,** 971–976 (2013).
126. Vétizou, M. *et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science* **350,** 1079–1084 (2015).
127. Charbonneau, M. R. *et al.* Sialylated milk oligosaccharides promote microbiota-dependent growth in models of infant undernutrition. *Cell* **164,** 859–871 (2016).
128. Cao, A. T. *et al.* Interleukin (IL)-21 promotes intestinal IgA response to microbiota. *Mucosal Immunol.* **8,** 1072–1082 (2015).
129. Kruglov, A. A. *et al.* Nonredundant function of soluble LTα3 produced by innate lymphoid cells in intestinal homeostasis. *Science* **342,** 1243–1246 (2013).
130. Sonnenberg, G. F., Monticelli, L. A., Elloso, M. M., Fouser, L. A. & Artis, D. CD4$^+$ lymphoid tissue-inducer cells promote innate immunity in the gut. *Immunity* **34,** 122–134 (2011).
131. Longman, R. S. *et al.* CX_3CR1^+ mononuclear phagocytes support colitis-associated innate lymphoid cell production of IL-22. *J. Exp. Med.* **211,** 1571–1583 (2014).
132. Cadwell, K. *et al.* Virus-plus-susceptibility gene interaction determines Crohn's disease gene *Atg16L1* phenotypes in intestine. *Cell* **141,** 1135–1145 (2010).
133. Kernbauer, E., Ding, Y. & Cadwell, K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature* **516,** 94–98 (2014).
134. Naik, S. *et al.* Commensal-dendritic-cell interaction specifies a unique protective skin immune signature. *Nature* **520,** 104–108 (2015).
135. Ichinohe, T. *et al.* Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc. Natl Acad. Sci. USA* **108,** 5354–5359 (2011).

Acknowledgements This work was supported by: grants from the Japan Agency for Medical Research and Development (AMED) and the Takeda Science Foundation (K.H.); US National Institutes of Health grant RO1DK103358 and the Howard Hughes Medical Institute (D.R.L.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: see go.nature.com/28j4zjb. Readers are welcome to comment on the online version of this paper at go.nature.com/28j4zjb. Correspondence should be addressed to K.H. (kenya@keio.jp) and D.L. (dan.littman@med.nyu.edu).

© 2016 Macmillan Publishers Limited. All rights reserved.

REVIEW

doi:10.1038/nature18849

Interactions between the microbiota and pathogenic bacteria in the gut

Andreas J. Bäumler[1] & Vanessa Sperandio[2,3]

The microbiome has an important role in human health. Changes in the microbiota can confer resistance to or promote infection by pathogenic bacteria. Antibiotics have a profound impact on the microbiota that alters the nutritional landscape of the gut and can lead to the expansion of pathogenic populations. Pathogenic bacteria exploit microbiota-derived sources of carbon and nitrogen as nutrients and regulatory signals to promote their own growth and virulence. By eliciting inflammation, these bacteria alter the intestinal environment and use unique systems for respiration and metal acquisition to drive their expansion. Unravelling the interactions between the microbiota, the host and pathogenic bacteria will produce strategies for manipulating the microbiota against infectious diseases.

Appreciation of the important role of the microbiota in human health and nutrition has grown steadily in the past decade. Initial studies focused on cataloguing the microbial species that comprise the microbiota and correlating the composition of the microbiota with the health or disease state of the host. The present period of renaissance has resulted in technologies and interdisciplinary research that are conducive to mechanistic studies and, in particular, those that focus on associations between the microbiota, the host and pathogenic bacteria. Exciting research is now starting to unravel how the composition of the microbiota can offer either resistance or assistance to invading pathogenic species. The majority of these studies were conducted in the gastrointestinal tract, in which associations between the host and microbes are of paramount importance. The gut microbiota of each individual is unique at the genus and species levels; however, it is more generally conserved at the phylum level, which is populated most prominently by Bacteroidetes and Firmicutes, followed by Proteobacteria and Actinobacteria. Host genetics, diet and environmental insults such as treatment with antibiotics alter the microbiota[1–4], which can lead to varying susceptibility to infectious diseases between individuals[5].

The microbiota can promote resistance to colonization by pathogenic species[6–9]. For instance, mice that are treated with antibiotics or that are bred in sterile environments (known as germ-free mice) are more susceptible to enteric pathogenic bacteria such as *Shigella flexneri, Citrobacter rodentium*, *Listeria monocytogenes* and *Salmonella enterica* serovar Typhimurium[10–13]. And some microbiotas can lead to the expansion or enhanced virulence of pathogenic populations[7]. A notable example concerns how differences in the composition of microbiotas determine the susceptibility of the mice to infection with *C. rodentium*: the transplantation of microbiotas from strains of mice that are susceptible to infection induced similar susceptibility in animals that were previously insusceptible, and the transplantation of microbiotas from resistant animals led to resistance to infection in previously susceptible animals[14,15]. Epidemiological surveys reinforce this idea. For example, differential susceptibility to infection with *Campylobacter jejuni* was shown to depend on the species composition of the microbiotas in a study of Swedish adults[16]. Individuals with a higher diversity within their microbiotas, and with an abundance of bacteria from the genera *Dorea* and *Coprococcus*, were significantly recalcitrant to *C. jejuni* infection compared with people who had low-diversity microbiotas and non-abundance of *Dorea* and *Coprococcus*.

The host's diet profoundly affects the composition of the microbiota, with repercussions for the physiology, immunity and susceptibility to infectious diseases of the host[17]. Dietary choices have been shown to affect colonization by enterohaemorrhagic *Escherichia coli* (EHEC) serotype O157:H7 and the severity and length of its resulting disease[18], and supplementation of the diet with phytonutrients promotes the expansion of beneficial Clostridia species that protect mice from colonization by *C. rodentium*[19].

The use of innovative technologies, in combination with more conventional approaches, is driving our understanding of the interactions between the microbiota, the host and pathogenic bacteria. The genetic tractability of several species of bacteria, as well as of their mammalian hosts (such as mice), allows for the mechanistic investigation of these relationships. The investigation of changes in the composition of microbiotas has been driven by next-generation sequencing, which also facilitated the analysis of transcriptomes. The growing power and finesse of metabolomics studies are quickly expanding our knowledge of the impact of both the microbiota and of pathogenic bacteria on the metabolic landscape of the gut. Here, we review advances in our understanding of the complex relationships that determine the severity and outcome of gastrointestinal infections. The majority of the mechanistic studies that investigate these interactions have been conducted in *S.* Typhimurium, EHEC and *Clostridium difficile*: therefore, these pathogenic organisms are covered more extensively than others in this Review.

Antibiotics

Antibiotics revolutionized medicine and were justifiably dubbed 'magic bullets' against bacterial infections. However, conventional antibiotics are generally bacteriostatic or bactericidal, which means that they indiscriminately kill or prevent the growth of both pathogenic and beneficial microbes. Antibiotics can alter the taxonomic, genomic and functional features of the microbiota, and their effects can be rapid and sometimes everlasting[20]. They can decrease the diversity of the microbiota, which compromises resistance to colonization by incoming pathogenic bacteria[20] — most notably leading to an expansion of *C. difficile* that can cause diarrhoea that leads to potentially fatal colitis[21].

C. difficile is a spore-forming bacterium that, on germination, colonizes the large intestine and causes colitis through the action of two toxins: TcdA and TcdB. The majority of *C. difficile* infections are nosocomial, but there has also been an increase in community-acquired infections, mainly due to the ubiquitous presence of *C. difficile* spores. *C. difficile* can

[1]Department of Medical Microbiology and Immunology, University of California, Davis, School of Medicine, Davis, California 95616, USA. [2]Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9048, USA. [3]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9038, USA.

© 2016 Macmillan Publishers Limited. All rights reserved.

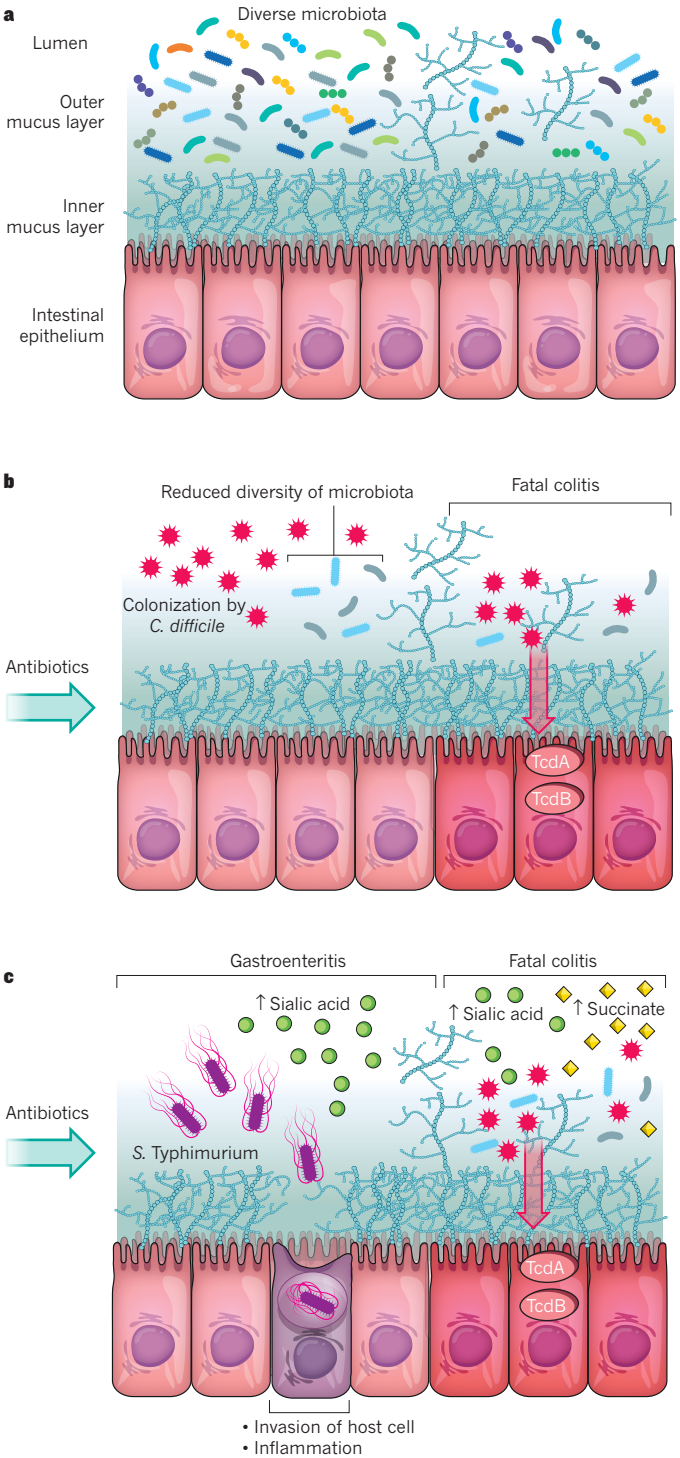


Figure 1 | The impact of antibiotics on the microbiota and the expansion of enteric pathogens. a, A diverse and non-disturbed microbiota confers resistance to colonization by enteric pathogens in the intestinal epithelium. **b**, Treatment with antibiotics decreases the diversity of the microbiota and leads to expansion of the *C. difficile* population. Toxins that are released from *C. difficile* (TcdA and TcdB) enter and damage the cells of the epithelium, which leads to inflammation (colitis) and cell death. **c**, Treatment with antibiotics also leads to an increase in the levels of free sialic acid (from the host) and succinate (from the microbiota) in the lumen of the intestine. Elevated sialic acid promotes the expansion of the *S.* Typhimurium population, which can lead to inflammation (gastroenteritis) if the bacterium invades the cells of the intestinal epithelium. Elevated levels of sialic acid and succinate further promote the expansion of the *C. difficile* population and the development of colitis and cell death.

colonize the mammalian intestine without causing disease, but one of the most important risk factors for colitis that is mediated by *C. difficile* is the use of antibiotics[21]. The antibiotics-mediated loss of resistance to colonization also allows colonization by *S.* Typhimurium and the development of disease[22]. Both *C. difficile* and *S.* Typhimurium catabolize sialic acid as a source of carbon in the lumen to promote their expansion[23]. They rely on saccharolytic members of the microbiota, such as *Bacteroides thetaiotaomicron*, to make this sugar freely available in the intestinal lumen. Treatment with antibiotics increases the abundance of host-derived free sialic acid as well as enhancing its release into the lumen by *B. thetaiotaomicron*, which promotes the expansion of the two pathogenic bacteria[23]. Antibiotic use also triggers production of the organic acid succinate, another microbiota-derived nutrient that confers a growth advantage to *C. difficile*. It is often present at a low concentration in the microbiotas of conventional mice, but its presence increases on treatment with antibiotics, which promotes a bloom of *C. difficile*[24] (Fig. 1).

Knowledge of how microbiota disruption affects the ability of bona fide or opportunistic pathogenic organisms to infect hosts is still in its infancy. However, two underlying themes converge: microbiota-induced changes in the metabolite landscape of the gut and inflammation.

Utilization of nutrients

Simple dietary sugars are absorbed in the small intestine, which means that they are unavailable as sources of carbon for the microbiota and pathogenic bacteria in the colon. The most abundant members of the microbiota are those that are able to utilize the undigested plant polysaccharides and host glycans that are present in the colon[25].

The gut epithelium is protected by a layer of mucus that is composed of proteins known as mucins that are rich in fucose, galactose, sialic acid, *N*-acetylgalactosamine, *N*-acetylglucosamine and mannose. These sugars are harvested by saccharolytic members of the microbiota, such as Bacteroidales in the gut, which makes them available to species within the microbiota that lack this capability[26]. However, pathogenic bacteria in the gut can also exploit the availability of these sugars to promote their own expansion. Several studies have used *B. thetaiotaomicron* as a model Bacteroides in which to investigate these syntrophic links. Sialic acid is a terminal sugar of some mucosal glycans, and *B. thetaiotaomicron* has sialidase activity but lacks the catabolic pathway for sialic-acid utilization. The bacterium therefore releases sialic acid to gain access to underlying glycans that it can use as a source of carbon. The sialic acid that *B. thetaiotaomicron* releases from the mucus can be catabolized by both *C. difficile* and *S.* Typhimurium, which provides them with a growth advantage[23]. The ability of the microbiota to use sialic acid therefore depends on the action of *B. thetaiotaomicron*, and mutants that lack sialidase fail to enhance the growth of these two pathogenic bacteria[23].

B. thetaiotaomicron also releases fucose from the mucus. It harbours multiple enzymes that can cleave fucose from host glycans, so its presence results in the high availability of fucose in the lumen of the gut[27–30]. This free fucose can also be used as a source of carbon by *S.* Typhimurium[23]. Importantly, *B. thetaiotaomicron* can promote the fucosylation of mucosal glycans when introduced into monoassociated germ-free mice[31,32].

The microbiota resides in the lumen and the outer mucus layer of the intestine. EHEC, however, aims to achieve a unique niche by closely adhering to the enterocytes of the intestinal epithelium. To achieve its goal, EHEC must successfully compete with the microbiota for nutrients. *B. thetaiotaomicron* does not need to compete with EHEC, however, because it can utilize polysaccharides; EHEC can only utilize monosaccharides and disaccharides[13,33]. EHEC's main competitors are commensal *E. coli*, which preferentially utilizes fucose as a source of carbon when growing in the mammalian intestine[13,33]. To circumvent this competition, EHEC utilizes other sources of sugar, such as galactose, the hexuranates, mannose and ribose, which commensal *E. coli* cannot catabolize optimally[33,34] (Fig. 2).

EHEC uses fucose as a signalling molecule with which to adjust its metabolism and to regulate the expression of its virulence repertoire in the lumen and the outer mucus layer of the colon[35]. It horizontally acquired

© 2016 Macmillan Publishers Limited. All rights reserved.

a pathogenicity island of genes that encode a fucose-sensing signalling-transduction system[35]. This system is unique to EHEC and to *C. rodentium*[35] (which is used extensively in mouse models as a surrogate for the human pathogen EHEC[36]). It is composed of the membrane-bound histidine sensor kinase FusK, which specifically autophosphorylates in response to fucose. FusK then transfers its phosphate to a response regulator called FusR, which is a transcription factor. Phosphorylation activates FusR, which represses the expression of the fucose utilization genes in EHEC, and helps EHEC to avoid the need to compete for this nutrient with commensal *E. coli*[35]. To prevent the unnecessary expenditure of energy by EHEC, FusR represses the genes that encode the EHEC virulence machinery, a syringe-like apparatus known as a type III secretion system (T3SS), which the bacterium uses to adhere itself to enterocytes and highjack the function of these host cells[35]. EHEC therefore uses fucose, a host-derived signal that is made available by the microbiota, to sense the environment of the intestinal lumen and to modulate its own metabolism and virulence.

To reach the lining of the epithelium, EHEC and *C. rodentium* produce mucinases[37], which cleave the protein backbone of mucin-type glycoproteins. Expression of these enzymes is increased by metabolites that are produced by *B. thetaiotaomicron*[38]. Because mucus is one of the main sources of sugar in the colon, where EHEC and *C. rodentium* colonize, obliteration of the mucus layer creates a nutrient-poor environment near the epithelium that is referred to as gluconeogenic. The colonization of mice by *B. thetaiotaomicron* therefore profoundly changes the metabolic landscape of the mouse gut because it raises the levels of organic acids such as succinate[24,38,39]. Moreover, several metabolites that indicate a gluconeogenic environment, such as lactate and glycerate, are also elevated[38]. EHEC and *C. rodentium* sense this gluconeogenic and succinate-rich environment through the transcriptional regulator Cra. On receiving the cue that they have reached the lining of the gut epithelium, these bacteria activate the expression of their T3SSs[38]. EHEC therefore exploits metabolic cues from *B. thetaiotaomicron*, and probably other members of the microbiota, to precisely programme its metabolism and virulence (Fig. 2).

Other pathogenic bacteria can also adjust their gene expression in the presence of microbiota-produced succinate. *C. difficile* induces a pathway that converts succinate to butyrate, which confers a growth advantage *in vivo*[24]. Populations of *C. difficile* mutants that are unable to convert succinate fail to expand in the gut in the presence of *B. thetaiotaomicron*[24].

Several short-chain fatty acids that are produced by the microbiota, are important determinants of interactions between the microbiota and pathogenic bacteria in the gut. The abundance and composition of short-chain fatty acids is distinct in each compartment of the intestine, and the ability to sense these differences might help pathogenic bacteria in niche recognition. The most abundant short-chain fatty acids in the gut are acetate, propionate and butyrate. *S.* Typhimurium preferably colonizes the ileum[40], which generally contains acetate at a concentration of 30 mM. This acetate concentration enhances the expression of the *S.* Typhimurium *Salmonella* pathogenicity island 1 (SPI-1)-encoded T3SS (T3SS-1), which is involved in the bacterium's invasion of the host. Conversely, 70 mM propionate and 20 mM butyrate, concentrations typical of the colon, suppress the expression of the T3SS-1 (ref. 41). Propionate and butyrate seem to affect the T3SS-1 regulatory cascade at various levels. However, the detailed mechanism of this regulation is yet to be unravelled. In EHEC, exposure to the levels of butyrate found in the colon increases the expression of the EHEC T3SS through post-transcriptional activation of the transcriptional regulator Lrp[42]. Exposure to the concentrations of acetate and propionate that are found in the small intestine does not significantly affect the virulence of EHEC.

Diet has a profound effect on the composition of the microbiota and the concentration of short-chain fatty acids in the gut[17]. A diet that is high in fibre results in the enhanced production of butyrate by the gut microbiota. That increases the host's expression of globotriaosylceramide, which is a receptor for the Shiga toxin that is produced by EHEC[18]. Shiga toxin can lead to the development of haemolytic uraemic syndrome (HUS) and is the cause of the morbidity and mortality associated with outbreaks of EHEC[43]. Consequently, animals that are fed a high-fibre diet are more susceptible to Shiga toxin than are those on a low-fibre diet and develop more severe disease[18]. Conversely, increased levels of microbiota-derived acetate protect animals from disease that is caused by the toxin. Certain species of *Bifidobacteria* contribute to higher levels of acetate in the gut, which helps to improve the barrier function of the intestinal epithelium and to prevent Shiga toxin from reaching the bloodstream[44].

Enteric pathogenic bacteria also use other nutrients to successfully overcome the microbiota's resistance to their colonization. Ethanolamine is abundant in the mammalian intestine[45]. It can be used as a source of

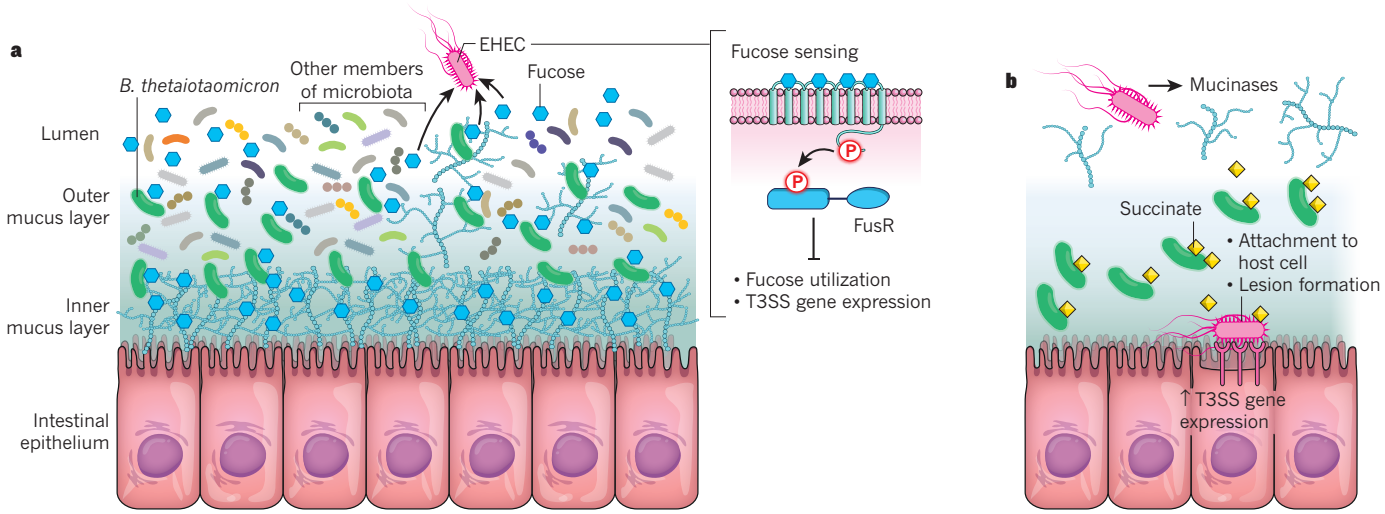


Figure 2 | Modulation of enterohaemorrhagic *E. coli* virulence through nutrients provided by the microbiota. a, The microbiota resides in the lumen and outer mucus layer of the intestine. The saccharolytic bacterium *Bacteroides thetaiotaomicron* is a prominent member of the microbiota. It can release fucose from the mucus and makes the sugar available to other bacteria. When EHEC senses fucose through the FusKR signalling system, it represses both its use of the sugar and the expression of genes that encode the T3SS, a protein-translocation apparatus that enables the bacterium to secrete effector proteins into host cells. This repression prevents EHEC from competing for fucose with commensal *E. coli* and from expending energy unnecessarily on T3SS expression. **b**, Metabolites that are provided by *B. thetaiotaomicron*, such as succinate, lead to an increase in the expression by EHEC of the enzyme mucinase, which obliterates the mucus layers of the intestine. EHEC is then able to reach the intestinal epithelium. *B. thetaiotaomicron* then begins to secrete succinate and other metabolites that are required for gluconeogenesis into the now nutrient-poor environment. The compounds are sensed by EHEC, which upregulates its expression of the T3SS to enable the bacterium to attach to the epithelial cells of the host intestine and form lesions that cause diarrhoea.

© 2016 Macmillan Publishers Limited. All rights reserved.

carbon and of nitrogen by a number of pathogenic species[46], and food-borne bacteria are particularly adept at using it. However, it cannot be metabolized by the majority of commensal species[47]. *S.* Typhimurium, EHEC and *L. monocytogenes* gain a growth advantage in the intestine through their ability to use this compound[45,48,49]. Ethanolamine is also used as a signal by EHEC and *S.* Typhimurium to activate the expression of virulence genes[50,51]. And *S.* Typhimurium uses hydrogen produced by the microbiota as an energy source to enhance its growth during the initial stage of infection[52].

The exploitation of microbiota-derived molecules as both nutrients and signals is crucial for the successful infection of the host by pathogenic bacteria. Although such organisms have clearly developed many strategies through which to circumvent the microbiota's resistance to colonization, and in many cases even employ its help, the microbiota pushes back, which creates an intense competition for resources. The ability of EHEC to colonize the intestine stems from differences in the sources of sugar that are used by EHEC and by commensal *E. coli*. For example, the presence of multiple strains of commensal *E. coli* with overlapping nutritional requirements interferes with the colonization of the mouse intestine by EHEC[53]. This study uses a streptomycin-treated mouse model of EHEC and three distinct commensal strains of *E. coli* to assess differential sugar requirements for the successful colonization of the intestines[53]. EHEC could colonize mice that were pre-colonized with any one of the commensal strains, but it could not colonize mice that were pre-colonized with all three strains[53]. EHEC has evolved to exploit distinct sources of sugar during colonization of the gut. It utilizes catabolic pathways for the hexuronates glucuronate and galacturonate and for sucrose that are not employed by commensal *E. coli* within the gut[33,53]. It can also metabolize several sugars simultaneously. The loss of multiple catabolic pathways has an additive effect on colonization. This phenomenon is not observed in commensal *E. coli*, however, which suggests that *E. coli* uses available sugars in a stepwise fashion[54]. EHEC therefore differs from commensal *E. coli* in metabolic strategy and the use of nutrients for the colonization of the mammalian intestine.

C. rodentium is outcompeted and then cleared from the mouse gut through a bloom in the population of commensal *E. coli* , which competes with *C. rodentium* for monosaccharides for nutrition[13]. By contrast, *C. rodentium* is not cleared by *B. thetaiotaomicron* in germ-free mice that are fed a diet that contains both monosaccharides, which can be used by Enterobacteriacae such as *C. rodentium*, and polysaccharides, which can be used by *Bacteroides*. However, when the mice are switched to a diet that consists only of monosaccharides, *B. thetaiotaomicron* and *C. rodentium* are forced to compete for sugars, and *B. thetaiotaomicron* outcompetes *C. rodentium*[13]. The ability of pathogenic bacteria to successfully compete with commensal species for nutrients is therefore important for their establishment in the gut.

Interception of signals from the microbiota and the host

The microbiota affects the risks and courses of enteric diseases. *Vibrio cholerae* is a major cause of explosive diarrhoea in which there is extensive disruption of the intestinal population of microbes. Metagenomic studies of the faecal microbiota of people with cholera in Bangladesh show that recovery is characterized by a certain microbiota signature. Reconstitution of this microbiota in germ-free mice restricts the infectivity of *V. cholerae*. Specifically, the presence of *Ruminococcus obeum* can hamper the colonization of the intestines by *V. cholerae* through the production of the furanone signal autoinducer-2, which causes the repression of several *V. cholerae* colonization factors[55].

Another example of the effect of microbiota-derived signals on host colonization is their use by EHEC in the colonization of its ruminal reservoir. EHEC exclusively colonizes the recto–anal junction of adult cattle. Through the sensor protein SdiA, EHEC detects acyl-homoserine lactone signals from the rumen microbiota, which it uses to reprogram itself to survive the acidic pH of the animal's stomachs and to successfully colonize the rectoanal junction[56].

As well as being able to directly detect signals that are derived from the microbiota, pathogenic bacteria can detect host-derived signals that have been modified by the microbiota to modulate their virulence. *V. cholerae* has a type VI secretion system (T6SS), which it uses to kill other bacteria. During its colonization of the intestine, *V. cholerae* comes in contact with the mucosal microbiota, which can affect the composition of bile acids in the intestine. For example, *Bifidobacterium bifidum* negatively regulates the T6SS activity of *V. cholerae* through the metabolic conversion of three bile acids (glycodeoxycholic acid, taurodeoxycholic acid and cholic acid) into the bile acid deoxycholic acid. Deoxycholic acid, but not its unmodified salts, decreases the expression of T6SS genes. This leads to a decrease in the killing of *E. coli* by *V. cholerae* owing to bile-acid conversion by other commensals, which decreases the activity of the T6SS[57].

Another microbiota-modified host signal that is detected by pathogenic bacteria is the neurotransmitter noradrenaline. The gut is highly innervated, and neurotransmitters are important signals in the gastrointestinal tract, where they modulate peristalsis, the flow of blood and the secretion of ions[58]. The microbiota affects the availability of neurotransmitters in the intestinal lumen, as well as their biosynthesis. For example, the microbiota induces biosynthesis of serotonin[59], and microbiota-derived enzymatic activities increase the levels of active noradrenaline in the gut lumen[60]. Noradrenaline is synthesized by the adrenergic neurons of the enteric nervous system[61] and it is inactivated by the host through conjugation with glucuronic acid (to produce a glucuronide). Microbiota-produced enzymes known as glucuronidases then deconjugate glucuronic acid from noradrenaline, which increases the amount of active noradrenaline in the lumen of the intestine[60]. Several pathogenic bacteria of the gut, including EHEC, *S.* Typhimurium and *V. parahaemolyticus*, sense noradrenaline to activate the expression of virulence genes[62–65]. Two adrenergic sensors have been identified in bacteria: the membrane-bound histidine kinases QseC and QseE[66,67]. QseC also detects the microbiota-produced signal autoinducer-3 (refs 64 and 66), so the sensing of signals from both the host and the microbiota converge at the level of a single receptor, a process known as inter-kingdom signalling.

Inflammation

Although diet and the composition of the microbiota heavily influence the availability of nutrients in the gut, the host also has an important part to play. A crucial driver of changes in the gut environment is the inflammatory response of the host. Intestinal inflammation in people is associated with an imbalance in the microbiota, known as dysbiosis, and is characterized by a reduced diversity of microbes, a reduced abundance of obligate anaerobic bacteria and an expansion of facultative anaerobic bacteria in the phylum Proteobacteria, mostly members of the family Enterobacteriaceae[68–73]. Similar changes in the composition of the gut microbiota are observed in mice with chemically induced colitis[74] and genetically induced colitis[75]. These changes in the structure of the microbiota probably reflect an altered nutritional environment that is created by the inflammatory response of the host.

The availability of nutrients in the large intestine is altered during inflammation through changes in the composition of mucous carbohydrates. Interleukin (IL)-22, a cytokine that is prominently induced in the intestinal mucosa when mice and rhesus macaques are infected with *S.* Typhimurium[76,77], stimulates the epithelial expression of galactoside 2-α-L-fucosyltransferase 2 and enhances the α(1,2)-fucosylation of mucus carbohydrates[78,79]. The gut microbiota can liberate fucose from mucus carbohydrates[23,80], which leads to the induction of genes for fucose utilization in *E. coli*[78]. Similarly, increased fucosylation of glycans is observed during *S.* Typhimurium-induced colitis in mice, which correlates with elevated synthesis of the proteins involved in fucose utilization[81]. Mucus fucosylation that is induced during infection with *C. rodentium* causes changes in the composition of the gut microbiota that help to protect the host from the expansion and epithelial translocation of the pathobiont *Enterococcus faecalis*[79].

Another driver of changes in the nutritional environment of the gut is the generation of reactive oxygen species and reactive nitrogen species during inflammation. Pro-inflammatory cytokines such as interferon-γ

© 2016 Macmillan Publishers Limited. All rights reserved.

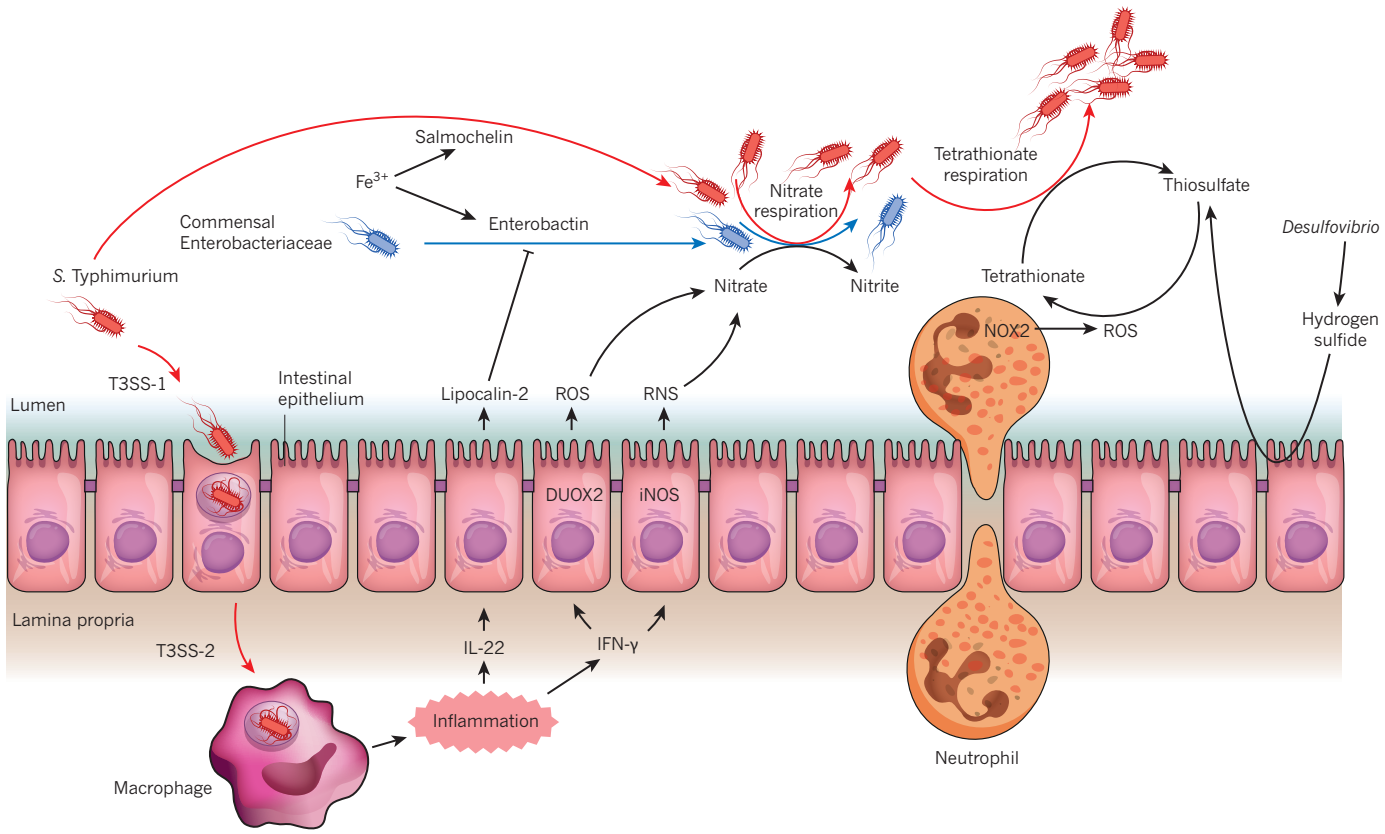


Figure 3 | The effect of intestinal inflammation on nutrient availability. *S.* Typhimurium uses its virulence factors (T3SS-1 and T3SS-2) to trigger intestinal inflammation. Cytokines that are released during inflammation, such as IL-22 and IFN-γ, trigger the release of antimicrobial molecules lipocalin-2, reactive oxygen species (ROS) and reactive nitrogen species (RNS) from the intestinal epithelium. Lipocalin-2 can block the growth of commensal Enterobacteriaceae that rely on the siderophore enterobactin for the acquisition of iron (Fe^{3+}). It does not bind to the *S.* Typhimurium siderophone salmochelin, however, which confers the bacterium with resistance to its effects on growth. RNS and ROS react to form nitrate, which drives the growth of Enterobacteriaceae through nitrate respiration. Microbiota-derived hydrogen sulfide is converted to thiosulfate by colonic epithelial cells. Neutrophils that migrate into the lumen of the intestine during inflammation generate ROS that convert endogenous sulfur compounds (thiosulfate) into an electron acceptor (tetrathionate) that further boosts the growth of *S.* Typhimurium through tetrathionate respiration.

(IFN-γ) activate dual oxidase 2 in the intestinal epithelium, which produces hydrogen peroxide[82]. Increased expression of *DUOX2*, the gene that encodes dual oxidase 2, in the intestinal mucosa of patients with Crohn's disease and ulcerative colitis correlates with an expansion of Proteobacteria in the gut microbiota[83]. IFN-γ also induces epithelial expression of the gene *Nos2* (ref. 84), which encodes inducible nitric oxide synthase, the enzyme that catalyses the production of nitric oxide from L-arginine[85]. As a result, the concentration of nitric oxide is elevated in gases from the colons of people with inflammatory bowel disease[86–88]. Although reactive oxygen and nitrogen species have antimicrobial activity, these radicals quickly form non-toxic compounds in the lumen of the gut as they diffuse away from the epithelium. For example, when they are generated during inflammation by host enzymes in the intestinal epithelium, these species react to form nitrate[89]. This by-product of inflammation is present at elevated concentrations in the intestines of mice with chemically induced colitis[90] (Fig. 3). Nitrate reductases, enzymes that are broadly conserved among the Enterobacteriaceae, couple the reduction of nitrate to energy-conserving electron transport systems for respiration, a process termed nitrate respiration. However, the genes that encode them are absent from the genomes of obligate anaerobic Clostridia or Bacteroidia[91]. Nitrate respiration drives the *Nos2*-dependent expansion of commensal *E. coli* in mice with chemically or genetically induced colitis, but not in animals without signs of intestinal inflammation[91]. Respiratory electron acceptors that are generated as a by-product of the host inflammatory response therefore create a niche in the lumen of the intestines that supports the uncontrolled expansion of commensal Enterobacteriaceae rather than of obligate anaerobic bacteria[91]. The resulting bloom in the inflamed intestine is one of the most consistent and robust ecological patterns that has been observed in the gut microbiota[92].

The creation of a niche for respiratory nutrients during inflammation is also an important driver of the strategies that pathogenic bacteria from the family Enterobacteriaceae use to invade the gut ecosystem. In the absence of inflammation or treatment with antibiotics, members of the gut microbiota occupy all available nutrient niches, which makes it very challenging for pathogenic Enterobacteriaceae to enter the community. One solution is for these bacteria to trigger intestinal inflammation, which would coerce the host into creating a fresh niche of respiratory nutrients that is suitable for its expansion — an approach that is used by *S.* Typhimurium[93]. On ingestion, *S.* Typhimurium uses T3SS-1 to invade the intestinal epithelium[94] and T3SS-2 to survive in the tissue of the host[95]. Both of these processes trigger acute intestinal inflammation in cattle and in mouse models of gastroenteritis[96–98] (Fig. 3). The inflammatory response of the host drives the expansion of *S.* Typhimurium in the lumen of the gut[99], which is required for the transmission of this pathogenic species to a new host through the faecal–oral route[100].

Although such expansion allows *S.* Typhimurium to side-step competition with obligate anaerobic Clostridia and Bacteroidia, this strategy forces the bacterium into battle with commensal Enterobacteriaceae over limited resources. For example, *S.* Typhimurium expands in the inflamed gut through nitrate respiration[101,102], which results in rivalry with commensal Enterobacteriaceae that pursue a similar strategy[91]. *S.* Typhimurium can gain an edge in this competition through its ability to utilize a broader range of inflammation-derived electron acceptors than its rivals. A source of one such electron acceptor is sulfate-reducing

© 2016 Macmillan Publishers Limited. All rights reserved.

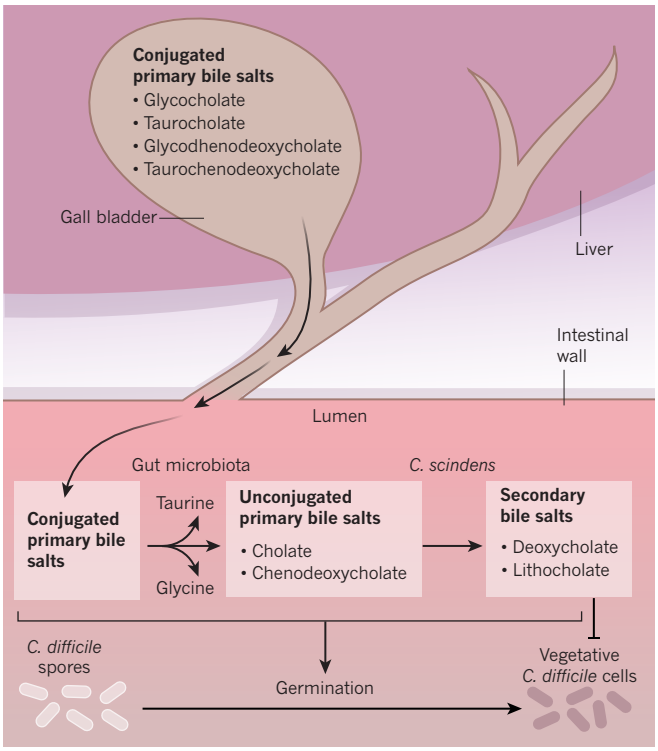
BOX 1

Microbiota interventions as therapeutic strategies to limit pathogen expansion

An imbalance in the gut microbiota might underlie many human diseases but, in most cases, the development of treatment options is still in its infancy. This could be in part because the mechanisms that lead to adverse effects in the host differ for each disease, which means that intervention strategies must be developed for each. The treatment options for antibiotic-induced dysbiosis are perhaps the most advanced, mainly because faecal microbiota transplantation can reverse this imbalance in the gut microbiota[129]. Nonetheless, the mechanisms through which treatment with antibiotics encourages an uncontrolled expansion of the obligate anaerobe *C. difficile* differ markedly from those that stimulate the growth of the facultative anaerobes Enterobacteriaceae, which has implications for the development of precision microbiome interventions.

Mice that are treated with streptomycin have a reduced abundance of members of the class Clostridia[130], which are credited with producing the lion's share of the short-chain fatty acid butyrate in the large intestine[131]. The resulting depletion of short-chain fatty acids drives an expansion of Enterobacteriaceae through mechanisms that are not fully resolved[44,132]. Depletion of Clostridia-derived butyrate affects the metabolism of enterocytes in the colon, which derive most of their energy by butyrate respiration[133]. The depletion of short-chain fatty acids also leads to a contraction in the pool of regulatory T cells in the colonic mucosa[134–136]. These changes in the host physiology increase the inflammatory tone of the mucosa, as indicated by the elevated expression of *Nos2*, the gene that encodes inducible nitric oxide synthase, and contributes to the expansion of commensal *E. coli* through nitrate generation[137]. Although other mechanisms probably contribute to the post-antibiotic expansion of certain populations of bacteria in the gut[126], the transfer of Clostridia, with their capacity for producing short-chain fatty acids, represents the most effective treatment for limiting the growth of *E. coli* in streptomycin-treated mice[138].

By contrast, the post-antibiotic expansion of the *C. difficile* population is driven by a depletion of secondary bile salts. The liver produces the primary bile salts cholate and chenodeoxycholate, which are conjugated to the amino acids taurine (to produce taurocholate and taurochenodeoxycholate) or glycine (to produce glycocholate and glycochenodeoxycholate) and then secreted into the gut. Bile salt hydrolases, enzymes that are produced by many members of the gut microbiota, remove the conjugated amino acid from the primary bile salt. *C. scindens* is one of a limited number of species of bacteria that can actively transport cholate and chenodeoxycholate into its cytosol, where these unconjugated primary bile salts are converted into the secondary bile salts deoxycholate and lithocholate, which are subsequently secreted into the extracellular environment[139] (Box Fig.). Although both primary and secondary bile salts induce the germination of *C. difficile* spores, only secondary bile salts efficiently prevent the growth of vegetative *C. difficile* cells[140]. By significantly reducing the abundance of species that are capable of producing deoxycholate and lithocholate, treatment with antibiotics causes a depletion of these secondary bile salts and promotes the expansion of vegetative *C. difficile* cells in the large intestine[141,142]. Faecal microbiota transplantation restores the production of secondary bile salts and therefore prevents the expansion of *C. difficile*[143]. Direct supplementation of the diet with secondary bile salts warrants caution because increased concentrations of bile salts have been linked to gastrointestinal cancers[144]. However, inoculation with only the secondary-bile-salt-producing *C. scindens* confers mice with resistance to *C. difficile* expansion following treatment with antibiotics[128]. This remarkable observation opens the door to novel precision microbiome interventions that aim to prevent or treat the colitis that is associated with *C. difficile* infection after antibiotic therapy.



species of *Desulfovibrio* from the microbiota, which release hydrogen sulfide, a compound that is converted to thiosulfate by the epithelium of the colon to avoid toxicity[103]. Deployment of the virulence factors of pathogenic bacteria leads to the recruitment of neutrophils to the intestinal mucosa, which is the histopathological hallmark of *S.* Typhimurium-induced gastroenteritis[96]. A fraction of these recruited neutrophils migrate into the lumen of the intestine — a diagnostic marker of inflammatory diarrhoea[104]. In the lumen, neutrophils help to protect the mucosa by engulfing bacteria in the vicinity of the epithelium[105], but reactive oxygen species that are generated by the phagocyte-produced NADPH oxidase 2 (also known as cytochrome b-245 heavy chain) convert thiosulfate into tetrathionate, a respiratory electron acceptor that supports the expansion of *S.* Typhimurium in the lumen of the inflamed gut[106] (Fig. 3). Although tetrathionate respiration is a characteristic of *Salmonella* serovars and has been used empirically in their isolation in clinical microbiology laboratories since 1923 (ref. 107), insights into the respiratory nutrient niche that *Salmonella* occupies suggest that this property is part of a strategy to edge out competing commensal Enterobacteriaceae in the inflamed gut[106].

The inflammatory response of the host also ignites competition between commensal and pathogenic Enterobacteriaceae over trace elements such as iron, which is less available during inflammation. IL-22 induces the release of the antimicrobial protein lipocalin-2 (also known as neutrophil gelatinase-associated lipocalin) from the epithelium in mice and rhesus macaques[108,109]. Lipocalin-2 reduces iron availability

© 2016 Macmillan Publishers Limited. All rights reserved.

by binding to enterobactin, a low-molecular-weight iron chelator (or siderophore) that is produced by Enterobacteriaceae[110,111]. To overcome this, *S.* Typhimurium and some commensal *E. coli* secrete a glycosylated derivative of enterobactin, termed salmochelin, which is not bound by lipocalin-2 (ref. 108). By producing salmochelin as well as two further siderophores that are not bound by lipocalin-2, yersiniobactin and aerobactin, the probiotic *E. coli* strain Nissle 1917 can limit the expansion of *S.* Typhimurium in the lumen of the inflamed gut[112]. Conversely, lipocalin-2 secretion by the epithelium generates an environment that enables *S.* Typhimurium to edge out commensal Enterobacteriaceae that depend solely on enterobactin for the acquisition of iron[109] (Fig. 3).

Through its limitation of iron availability, intestinal inflammation also sets the stage for battles between Enterobacteriaceae that use protein-based toxins known as colicins[113] that affect a narrow range of hosts. Iron limitation induces the synthesis of siderophore receptor proteins for the bacterial outer membrane[113], which also commonly serve as receptors for colicins[114–116]. Expression of a siderophore receptor protein termed the colicin I receptor (CirA) confers commensal *E. coli* with sensitivity to colicin Ib produced by *S.* Typhimurium[113]. The respiratory nutrient niche that is generated by the inflammatory response of the host is therefore a battleground on which commensal and pathogenic Enterobacteriaceae struggle for dominance using a diverse arsenal of nutritional and antimicrobial strategies.

Perspective and the future

The study of the microbiome began more than a century ago. equencing of 16S rRNA genes provided the first insights into the taxonomic composition of microbial communities. Later, sequencing of the complete metagenome of microbial communities provided a more detailed insight into the full genetic capacity of such a community. The use of germ-free animals, either alone or in combination with emerging technologies such as laser-capture microdissection and transcriptomics, enabled mechanistic studies of the associations between the microbiota, the host and pathogenic bacteria[117]. Multi-taxon insertion sequencing now allows researchers to investigate both the assembly and the shared and strain-specific dietary requirements of communities of microbes, and it has also facilitated the informed manipulation of such communities through diet[118]. The development of quantitative imaging technologies has provided insight into the localization of microbes within the gastrointestinal tract, and it has also enabled studies on the proximity of and interactions between microbes[119]. The increasing refinement and power of metabolomics, imaging mass spectrometry and three-dimensional mapping of mass-spectrometry data provide a high-resolution image of the complex chemistry landscape of the interactions between microbes and the host, which sets the stage for manipulating this chemistry to prevent or treat infectious diseases[24,38,120–127]. A marriage of metagenomics and mathematical modelling promises to enhance the precision of microbiome reconstitution, which has proven successful for tackling *C. difficile* infections in mice[128]. In these exciting times, the expansion of multidisciplinary research is rapidly generating new technologies and mechanistic insights into interactions between the microbiota, the host and pathogenic bacteria (Box 1). ■

Received 4 September 2015; accepted 22 April 2016.

1. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308,** 1635–1638 (2005).
2. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104,** 13780–13785 (2007).
3. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
4. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124,** 837–848 (2006).
5. Yurist-Doutsch, S., Arrieta, M. C., Vogt, S. L. & Finlay, B. B. Gastrointestinal microbiota-mediated control of enteric pathogens. *Annu. Rev. Genet.* **48,** 361–382 (2014).
6. Sassone-Corsi, M. & Raffatellu, M. No vacancy: how beneficial microbes cooperate with immunity to provide colonization resistance to pathogens. *J. Immunol.* **194,** 4081–4087 (2015).
7. Cameron, E. A. & Sperandio, V. Frenemies: signaling and nutritional integration in pathogen–microbiota–host interactions. *Cell Host Microbe* **18,** 275–284 (2015).
8. Pacheco, A. R. & Sperandio, V. Enteric pathogens exploit the microbiota-generated nutritional environment of the gut. *Microbiol. Spectr.* **3,** MBP-0001-2014 (2015).
9. Bohnhoff, M., Drake, B. L. & Miller, C. P. Effect of streptomycin on susceptibility of intestinal tract to experimental *Salmonella* infection. *Proc. Soc. Exp. Biol. Med.* **86,** 132–137 (1954).
10. Ferreira, R. B. *et al.* The intestinal microbiota plays a role in *Salmonella*-induced colitis independent of pathogen colonization. *PLoS ONE* **6,** e20338 (2011).
11. Sprinz, H. *et al.* The response of the germfree guinea pig to oral bacterial challenge with *Escherichia coli* and *Shigella flexneri*. *Am. J. Pathol.* **39,** 681–695 (1961).
12. Zachar, Z. & Savage, D. C. Microbial interference and colonization of the murine gastrointestinal tract by *Listeria monocytogenes*. *Infect. Immun.* **23,** 168–174 (1979).
13. Kamada, N. *et al.* Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science* **336,** 1325–1329 (2012).
This study showed that enteric pathogens use virulence genes to compete with the microbiota for nutrients in the gut.
14. Ghosh, S. *et al.* Colonic microbiota alters host susceptibility to infectious colitis by modulating inflammation, redox status, and ion transporter gene expression. *Am. J. Physiol. Gastrointest. Liver Physiol.* **301,** G39–G49 (2011).
15. Willing, B. P., Vacharaksa, A., Croxen, M., Thanachayanont, T. & Finlay, B. B. Altering host resistance to infections through microbial transplantation. *PLoS ONE* **6,** e26988 (2011).
16. Kampmann, C., Dicksved, J., Engstrand, L. & Rautelin, H. Composition of human faecal microbiota in resistance to *Campylobacter* infection. *Clin. Microbiol. Infect.* **22,** 61.e1–61.e8 (2016).
17. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474,** 327–336 (2011).
18. Zumbrun, S. D. *et al.* Dietary choice affects Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 colonization and disease. *Proc. Natl Acad. Sci. USA* **110,** E2126–E2133 (2013).
19. Wlodarska, M., Willing, B. P., Bravo, D. M. & Finlay, B. B. Phytonutrient diet supplementation promotes beneficial Clostridia species and intestinal mucus secretion resulting in protection against enteric infection. *Sci. Rep.* **5,** 9253 (2015).
20. Modi, S. R., Collins, J. J. & Relman, D. A. Antibiotics and the gut microbiota. *J. Clin. Invest.* **124,** 4212–4218 (2014).
21. Leffler, D. A. & Lamont, J. T. *Clostridium difficile* infection. *N. Engl. J. Med.* **373,** 287–288 (2015).
22. Pavia, A. T. *et al.* Epidemiologic evidence that prior antimicrobial exposure decreases resistance to infection by antimicrobial-sensitive *Salmonella*. *J. Infect. Dis.* **161,** 255–260 (1990).
23. Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502,** 96–99 (2013).
This study showed that treatment with antibiotics enhances the abundance of host sialic acid that can be harvested by the microbiota, which promotes the expansion of enteric pathogens.
24. Ferreyra, J. A. *et al.* Gut microbiota-produced succinate promotes *C. difficile* infection after antibiotic treatment or motility disturbance. *Cell Host Microbe* **16,** 770–777 (2014).
25. Ferreyra, J. A., Ng, K. M. & Sonnenburg, J. L. The enteric two-step: nutritional strategies of bacterial pathogens within the gut. *Cell. Microbiol.* **16,** 993–1003 (2014).
26. Rakoff-Nahoum, S., Coyne, M. J. & Comstock, L. E. An ecological network of polysaccharide utilization among human intestinal symbionts. *Curr. Biol.* **24,** 40–49 (2014).
27. Alverdy, J., Chi, H. S. & Sheldon, G. F. The effect of parenteral nutrition on gastrointestinal immunity. The importance of enteral stimulation. *Ann. Surg.* **202,** 681–684 (1985).
28. Fischbach, M. A. & Sonnenburg, J. L. Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* **10,** 336–347 (2011).
29. Chow, W. L. & Lee, Y. K. Free fucose is a danger signal to human intestinal epithelial cells. *Br. J. Nutr.* **99,** 449–454 (2008).
30. Bourlioux, P., Koletzko, B., Guarner, F. & Braesco, V. The intestine and its microflora are partners for the protection of the host: report on the Danone Symposium "The Intelligent Intestine," held in Paris, June 14, 2002. *Am. J. Clin. Nutr.* **78,** 675–683 (2003).
31. Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I. A model of host–microbial interactions in an open mammalian ecosystem. *Science* **273,** 1380–1383 (1996).
32. Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc. Natl Acad. Sci. USA* **96,** 9833–9838 (1999).
33. Fabich, A. J. *et al.* Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infect. Immun.* **76,** 1143–1152 (2008).
34. Autieri, S. M. *et al.* L-fucose stimulates utilization of D-ribose by *Escherichia coli* MG1655 *ΔfucAO* and *E. coli* Nissle 1917 *ΔfucAO* mutants in the mouse intestine and in M9 minimal medium. *Infect. Immun.* **75,** 5465–5475 (2007).
35. Pacheco, A. R. *et al.* Fucose sensing regulates bacterial intestinal colonization. *Nature* **492,** 113–117 (2012).
This paper showed that sugars from the mucus that are released by the

© 2016 Macmillan Publishers Limited. All rights reserved.

microbiota can be perceived as signals by enteric pathogens to regulate the expression of virulence genes.
36. Schauer, D. B. & Falkow, S. Attaching and effacing locus of a *Citrobacter freundii* biotype that causes transmissible murine colonic hyperplasia. *Infect. Immun.* **61,** 2486–2492 (1993).
37. Szabady, R. L., Lokuta, M. A., Walters, K. B., Huttenlocher, A. & Welch, R. A. Modulation of neutrophil function by a secreted mucinase of *Escherichia coli* O157:H7. *PLoS Pathog.* **5,** e1000320 (2009).
38. Curtis, M. M. *et al.* The gut commensal *Bacteroides thetaiotaomicron* exacerbates enteric infection through modification of the metabolic landscape. *Cell Host Microbe* **16,** 759–769 (2014).
This study revealed that metabolites that are produced by the microbiota can be exploited as cues to increase the virulence of enteric pathogens.
39. Macy, J. M., Ljungdahl, L. G. & Gottschalk, G. Pathway of succinate and propionate formation in *Bacteroides fragilis*. *J. Bacteriol.* **134,** 84–91 (1978).
40. Carter, P. B. & Collins, F. M. The route of enteric infection in normal mice. *J. Exp. Med.* **139**, 1189–1203 (1974).
41. Lawhon, S. D., Maurer, R., Suyemoto, M. & Altier, C. Intestinal short-chain fatty acids alter *Salmonella typhimurium* invasion gene expression and virulence through BarA/SirA. *Mol. Microbiol.* **46,** 1451–1464 (2002).
42. Takao, M., Yen, H. & Tobe, T. LeuO enhances butyrate-induced virulence expression through a positive regulatory loop in enterohaemorrhagic *Escherichia coli*. *Mol. Microbiol.* **93,** 1302–1313 (2014).
43. Karmali, M. A., Petric, M., Lim, C., Fleming, P. C. & Steele, B. T. *Escherichia coli* cytotoxin, haemolytic-uraemic syndrome, and haemorrhagic colitis. *Lancet* **2,** 1299–1300 (1983).
44. Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469,** 543–547 (2011).
This study showed that the acetate that is produced by certain probiotic bacteria can enhance the barrier function of the gut epithelium to protect the host from enteric infections.
45. Bertin, Y. *et al.* Enterohaemorrhagic *Escherichia coli* gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. *Environ. Microbiol.* **13,** 365–377 (2011).
46. Garsin, D. A. Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nature Rev. Microbiol.* **8,** 290–295 (2010).
47. Korbel, J. O. *et al.* Systematic association of genes to phenotypes by genome and literature mining. *PLoS Biol.* **3,** e134 (2005).
48. Thiennimitr, P. *et al.* Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. *Proc. Natl Acad. Sci. USA* **108,** 17480–17485 (2011).
49. Joseph, B. *et al.* Identification of *Listeria monocytogenes* genes contributing to intracellular replication by expression profiling and mutant screening. *J. Bacteriol.* **188,** 556–568 (2006).
50. Kendall, M. M., Gruber, C. C., Parker, C. T. & Sperandio, V. Ethanolamine controls expression of genes encoding components involved in interkingdom signaling and virulence in enterohemorrhagic *Escherichia coli* O157:H7. *mBio* **3,** 00050-12 (2012).
51. Anderson, C. J., Clark, D. E., Adli, M. & Kendall, M. M. Ethanolamine signaling promotes *Salmonella* niche recognition and adaptation during infection. *PLoS Pathog.* **11,** e1005278 (2015).
52. Maier, L. *et al.* Microbiota-derived hydrogen fuels *Salmonella* Typhimurium invasion of the gut ecosystem. *Cell Host Microbe* **14,** 641–651 (2013).
53. Maltby, R., Leatham-Jensen, M. P., Gibson, T., Cohen, P. S. & Conway, T. Nutritional basis for colonization resistance by human commensal *Escherichia coli* strains HS and Nissle 1917 against *E. coli* O157:H7 in the mouse intestine. *PLoS ONE* **8,** e53957 (2013).
54. Fabich, A. J. *et al.* Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine. *Infect. Immun.* **76,** 1143–1152 (2008).
55. Hsiao, A. *et al.* Members of the human gut microbiota involved in recovery from *Vibrio cholerae* infection. *Nature* **515,** 423–426 (2014).
56. Hughes, D. T. *et al.* Chemical sensing in mammalian host–bacterial commensal associations. *Proc. Natl Acad. Sci. USA* **107,** 9831–9836 (2010).
57. Bachmann, V. *et al.* Bile salts modulate the mucin-activated type VI secretion system of pandemic *Vibrio cholerae*. *PLoS Negl. Trop. Dis.* **9,** e0004031 (2015).
58. Hörger, S., Schultheiss, G. & Diener, M. Segment-specific effects of epinephrine on ion transport in the colon of the rat. *Am. J. Physiol.* **275,** G1367–G1376 (1998).
59. Yano, J. M. *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161,** 264–276 (2015).
60. Asano, Y. *et al.* Critical role of gut microbiota in the production of biologically active, free catecholamines in the gut lumen of mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* **303,** G1288–G1295 (2012).
61. Furness, J. B. Types of neurons in the enteric nervous system. *J. Auton. Nerv. Syst.* **81,** 87–96 (2000).
62. Curtis, M. M. & Sperandio, V. A complex relationship: the interaction among symbiotic microbes, invading pathogens, and their mammalian host. *Mucosal Immunol.* **4,** 133–138 (2011).
63. Moreira, C. G., Weinshenker, D. & Sperandio, V. QseC mediates *Salmonella enterica* serovar Typhimurium virulence *in vitro* and *in vivo*. *Infect. Immun.* **78,** 914–926 (2010).
64. Sperandio, V., Torres, A. G., Jarvis, B., Nataro, J. P. & Kaper, J. B. Bacteria–host communication: the language of hormones. *Proc. Natl Acad. Sci. USA* **100,** 8951–8956 (2003).
This paper describes how signals from the microbiota and host converge to enhance virulence in enteric pathogens.
65. Nakano, M., Takahashi, A., Sakai, Y. & Nakaya, Y. Modulation of pathogenicity with norepinephrine related to the type III secretion system of *Vibrio parahaemolyticus*. *J. Infect. Dis.* **195,** 1353–1360 (2007).
66. Clarke, M. B., Hughes, D. T., Zhu, C., Boedeker, E. C. & Sperandio, V. The QseC sensor kinase: a bacterial adrenergic receptor. *Proc. Natl Acad. Sci. USA* **103,** 10420–10425 (2006).
67. Reading, N. C., Rasko, D. A., Torres, A. G. & Sperandio, V. The two-component system QseEF and the membrane protein QseG link adrenergic and stress sensing to bacterial pathogenesis. *Proc. Natl Acad. Sci. USA* **106,** 5889–5894 (2009).
68. Seksik, P. *et al.* Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* **52,** 237–242 (2003).
69. Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F. & Veldhuyzen van Zanten, S. J. Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J. Clin. Microbiol.* **44,** 4136–4141 (2006).
70. Baumgart, M. *et al.* Culture independent analysis of ileal mucosa reveals a selective increase in invasive *Escherichia coli* of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *ISME J.* **1,** 403–418 (2007).
71. Walker, A. W. *et al.* High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol.* **11,** 7 (2011).
72. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15,** 382–392 (2014).
73. Chiodini, R. J. *et al.* Microbial population differentials between mucosal and submucosal intestinal tissues in advanced Crohn's disease of the ileum. *PLoS ONE* **10,** e0134382 (2015).
74. Lupp, C. *et al.* Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* **2,** 119–129; corrigendum **2,** 204 (2007).
This paper reported that the induction of inflammation in the host disrupts the microbiota and promotes a bloom of Enterobacteriaceae.
75. Garrett, W. S. *et al.* Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8,** 292–300 (2010).
76. Raffatellu, M. *et al.* Simian immunodeficiency virus-induced mucosal interleukin-17 deficiency promotes *Salmonella* dissemination from the gut. *Nature Med.* **14,** 421–428 (2008).
77. Godinez, I. *et al.* T cells help to amplify inflammatory responses induced by *Salmonella enterica* serotype Typhimurium in the intestinal mucosa. *Infect. Immun.* **76,** 2008–2017 (2008).
78. Pickard, J. M. *et al.* Rapid fucosylation of intestinal epithelium sustains host–commensal symbiosis in sickness. *Nature* **514,** 638–641 (2014).
This paper showed that cytokines that are induced by enteric infection promote fucosylation in the host.
79. Pham, T. A. *et al.* Epithelial IL-22RA1-mediated fucosylation promotes intestinal colonization resistance to an opportunistic pathogen. *Cell Host Microbe* **16,** 504–516 (2014).
80. Sonnenburg, J. L. *et al.* Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307,** 1955–1959 (2005).
81. Ansong, C. *et al.* Studying Salmonellae and Yersiniae host–pathogen interactions using integrated 'omics and modeling. *Curr. Top. Microbiol. Immunol.* **363,** 21–41 (2013).
82. Harper, R. W. *et al.* Differential regulation of dual NADPH oxidases/peroxidases, Duox1 and Duox2, by Th1 and Th2 cytokines in respiratory tract epithelium. *FEBS Lett.* **579,** 4911–4917 (2005).
83. Haberman, Y. *et al.* Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J. Clin. Invest.* **124,** 3617–3633 (2014).
84. Salzman, A. *et al.* Induction and activity of nitric oxide synthase in cultured human intestinal epithelial monolayers. *Am. J. Physiol.* **270,** G565–G573 (1996).
85. Palmer, R. M., Rees, D. D., Ashton, D. S. & Moncada, S. L-arginine is the physiological precursor for the formation of nitric oxide in endothelium-dependent relaxation. *Biochem. Biophys. Res. Commun.* **153,** 1251–1256 (1988).
86. Lundberg, J. O., Weitzberg, E., Lundberg, J. M. & Alving, K. Intragastric nitric oxide production in humans: measurements in expelled air. *Gut* **35,** 1543–1546 (1994).
87. Singer, I. I. *et al.* Expression of inducible nitric oxide synthase and nitrotyrosine in colonic epithelium in inflammatory bowel disease. *Gastroenterology* **111,** 871–885 (1996).
88. Enocksson, A., Lundberg, J., Weitzberg, E., Norrby-Teglund, A. & Svenungsson, B. Rectal nitric oxide gas and stool cytokine levels during the course of infectious gastroenteritis. *Clin. Diagn. Lab. Immunol.* **11,** 250–254 (2004).
89. Bai, P. *et al.* Protein tyrosine nitration and poly(ADP-ribose) polymerase activation in *N*-methyl-*N*-nitro-*N*-nitrosoguanidine-treated thymocytes: implication for cytotoxicity. *Toxicol. Lett.* **170,** 203–213 (2007).
90. Dudhgaonkar, S. P., Tandan, S. K., Kumar, D., Raviprakash, V. & Kataria, M. Influence of simultaneous inhibition of cyclooxygenase-2 and inducible nitric oxide synthase in experimental colitis in rats. *Inflammopharmacology* **15,** 188–195 (2007).
91. Winter, S. E. *et al.* Host-derived nitrate boosts growth of *E. coli* in the inflamed gut. *Science* **339**, 708–711 (2013).

© 2016 Macmillan Publishers Limited. All rights reserved.

This study revealed that nitrate respiration drives the expansion of commensal *E. coli* during colitis.
92. Winter, S. E. & Baumler, A. J. Why related bacterial species bloom simultaneously in the gut: principles underlying the 'like will to like' concept. *Cell. Microbiol.* **16,** 179–184 (2014).
93. Rivera-Chávez, F. & Baumler, A. J. The pyromaniac inside you: *Salmonella* metabolism in the host gut. *Annu. Rev. Microbiol.* **69,** 31–48 (2015).
94. Galán, J. E. & Curtiss, R., III. Cloning and molecular characterization of genes whose products allow *Salmonella typhimurium* to penetrate tissue culture cells. *Proc. Natl Acad. Sci. USA* **86,** 6383–6387 (1989).
95. Hensel, M. *et al.* Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269,** 400–403 (1995).
96. Tsolis, R. M., Adams, L. G., Ficht, T. A. & Baumler, A. J. Contribution of *Salmonella typhimurium* virulence factors to diarrheal disease in calves. *Infect. Immun.* **67,** 4879–4885 (1999).
97. Barthel, M. *et al.* Pretreatment of mice with streptomycin provides a *Salmonella enterica* serovar Typhimurium colitis model that allows analysis of both pathogen and host. *Infect. Immun.* **71,** 2839–2858 (2003).
98. Coburn, B., Li, Y., Owen, D., Vallance, B. A. & Finlay, B. B. *Salmonella enterica* serovar Typhimurium pathogenicity island 2 is necessary for complete virulence in a mouse model of infectious enterocolitis. *Infect. Immun.* **73,** 3219–3227 (2005).
99. Stecher, B. *et al. Salmonella enterica* serovar Typhimurium exploits inflammation to compete with the intestinal microbiota. *PLoS Biol.* **5,** 2177–2189 (2007).
100. Lawley, T. D. *et al.* Host transmission of *Salmonella enterica* serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. *Infect. Immun.* **76,** 403–416 (2008).
101. Lopez, C. A. *et al.* Phage-mediated acquisition of a type III secreted effector protein boosts growth of *Salmonella* by nitrate respiration. *mBio* **3,** e00143-12 (2012).
102. Lopez, C. A., Rivera-Chavez, F., Byndloss, M. X. & Baumler, A. J. The periplasmic nitrate reductase NapABC supports luminal growth of *Salmonella enterica* serovar Typhimurium during colitis. *Infect. Immun.* **83,** 3470–3478 (2015).
103. Levitt, M. D., Furne, J., Springfield, J., Suarez, F. & DeMaster, E. Detoxification of hydrogen sulfide and methanethiol in the cecal mucosa. *J. Clin. Invest.* **104,** 1107–1114 (1999).
104. Harris, J. C., Dupont, H. L. & Hornick, R. B. Fecal leukocytes in diarrheal illness. *Ann. Intern. Med.* **76,** 697–703 (1972).
105. Loetscher, Y. *et al.* Salmonella transiently reside in luminal neutrophils in the inflamed gut. *PLoS ONE* **7,** e34812 (2012).
106. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* **467,** 426–429 (2010).
This paper showed that tetrathionate is a unique electron acceptor that is induced through inflammation that promotes expansion of *S.* Typhimurium.
107. Muller, H. J. Partial list of biological institutes and biologists doing experimental work in Russia at the present time. *Science* **57,** 472–473 (1923).
108. Raffatellu, M. *et al.* Lipocalin-2 resistance confers an advantage to *Salmonella enterica* serotype Typhimurium for growth and survival in the inflamed intestine. *Cell Host Microbe* **5,** 476–486 (2009).
109. Behnsen, J. *et al.* The cytokine IL-22 promotes pathogen colonization by suppressing related commensal bacteria. *Immunity* **40,** 262–273 (2014).
110. Goetz, D. H. *et al.* The neutrophil lipocalin NGAL is a bacteriostatic agent that interferes with siderophore-mediated iron acquisition. *Mol. Cell* **10,** 1033–1043 (2002).
111. Flo, T. H. *et al.* Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. *Nature* **432,** 917–921 (2004).
112. Deriu, E. *et al.* Probiotic bacteria reduce *Salmonella typhimurium* intestinal colonization by competing for iron. *Cell Host Microbe* **14,** 26–37 (2013).
This paper revealed that probiotic strains of *E. coli* restrict infection of the gut with *S.* Typhimurium by competing for sources of iron.
113. Cascales, E. *et al.* Colicin biology. *Microbiol. Mol. Biol. Rev.* **71,** 158–229 (2007).
114. Guterman, S. K. Colicin B: mode of action and inhibition by enterochelin. *J. Bacteriol.* **114,** 1217–1224 (1973).
115. Cardelli, J. & Konisky, J. Isolation and characterization of an *Escherichia coli* mutant tolerant to colicins Ia and Ib. *J. Bacteriol.* **119,** 379–385 (1974).
116. Patzer, S. I., Baquero, M. R., Bravo, D., Moreno, F. & Hantke, K. The colicin G, H and X determinants encode microcins M and H47, which might utilize the catecholate siderophore receptors FepA, Cir, Fiu and IroN. *Microbiology* **149,** 2557–2570 (2003).
117. Hooper, L. V. *et al.* Molecular analysis of commensal host–microbial relationships in the intestine. *Science* **291,** 881–884 (2001).
This was the first report to show that the microbiota can change the expression of mammalian genes to modulate the immune system of the host.
118. Wu, M. *et al.* Genetic determinants of *in vivo* fitness and diet responsiveness in multiple human gut *Bacteroides*. *Science* **350,** aac5992 (2015).
119. Earle, K. A. *et al.* Quantitative imaging of gut microbiota spatial organization. *Cell Host Microbe* **18,** 478–488 (2015).
120. Bouslimani, A. *et al.* Molecular cartography of the human skin surface in 3D. *Proc. Natl Acad. Sci. USA* **112,** E2120–E2129 (2015).
121. Marcobal, A. *et al.* A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME J.* **7,** 1933–1943 (2013).
122. Lau, W., Fischbach, M. A., Osbourn, A. & Sattely, E. S. Key applications of plant metabolic engineering. *PLoS Biol.* **12,** e1001879 (2014).
123. Marcobal, A. *et al.* Metabolome progression during early gut microbial colonization of gnotobiotic mice. *Sci. Rep.* **5,** 11589 (2015).
124. Rath, C. M. *et al.* Molecular analysis of model gut microbiotas by imaging mass spectrometry and nanodesorption electrospray ionization reveals dietary metabolite transformations. *Anal. Chem.* **84,** 9259–9267 (2012).
This study used imaging mass spectrometry to map differences in the gut metabolic landscape that are promoted by the microbiota.
125. Dorrestein, P. C., Mazmanian, S. K. & Knight, R. Finding the missing links among metabolites, microbes, and the host. *Immunity* **40,** 824–832 (2014).
126. Antunes, L. C. *et al.* Antivirulence activity of the human gut metabolome. *mBio* **5,** e01183-14 (2014).
127. Antunes, L. C. *et al.* Impact of *Salmonella* infection on host hormone metabolism revealed by metabolomics. *Infect. Immun.* **79,** 1759–1769 (2011).
128. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517,** 205–208 (2015).
This paper showed that the introduction of a single species of bacteria that can produce secondary bile salts confers the host with resistance to the expansion of *C. difficile* after treatment with antibiotics.
129. Koenigsknecht, M. J. & Young, V. B. Faecal microbiota transplantation for the treatment of recurrent *Clostridium difficile* infection: current promise and future needs. *Curr. Opin. Gastroenterol.* **29,** 628–632 (2013).
130. Sekirov, I. *et al.* Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infect. Immun.* **76,** 4726–4736 (2008).
131. Louis, E., Libioulle, C., Reenaers, C., Belaiche, J. & Georges, M. Genomics of inflammatory bowel diseases: basis for a new molecular classification and new therapeutic strategies of these diseases [in French]. *Rev. Med. Liege* **64** S1, 24–28 (2009).
132. Meynell, G. G. Antibacterial mechanisms of the mouse gut. II. The role of Eh and volatile fatty acids in the normal gut. *Br. J. Exp. Pathol.* **44,** 209–219 (1963).
133. Donohoe, D. R., Wali, A., Brylawski, B. P. & Bultman, S. J. Microbial regulation of glucose metabolism and cell-cycle progression in mammalian colonocytes. *PLoS ONE* **7,** e46589 (2012).
134. Atarashi, K., Umesaki, Y. & Honda, K. Microbiotal influence on T cell subset development. *Semin. Immunol.* **23,** 146–153 (2011).
135. Atarashi, K. *et al.* T_{reg} induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).
136. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504,** 446–450 (2013).
137. Spees, A. M. *et al.* Streptomycin-induced inflammation enhances *Escherichia coli* gut colonization through nitrate respiration. *mBio* **4,** e00430-13 (2013).
138. Itoh, K. & Freter, R. Control of *Escherichia coli* populations by a combination of indigenous clostridia and lactobacilli in gnotobiotic mice and continuous-flow cultures. *Infect. Immun.* **57,** 559–565 (1989).
139. Wells, J. E. & Hylemon, P. B. Identification and characterization of a bile acid 7α-dehydroxylation operon in *Clostridium* sp. strain TO-931, a highly active 7α-dehydroxylating strain isolated from human feces. *Appl. Environ. Microbiol.* **66,** 1107–1113 (2000).
140. Wilson, K. H. Efficiency of various bile salt preparations for stimulation of *Clostridium difficile* spore germination. *J. Clin. Microbiol.* **18,** 1017–1019 (1983).
141. Sorg, J. A. & Sonenshein, A. L. Bile salts and glycine as cogerminants for *Clostridium difficile* spores. *J. Bacteriol.* **190,** 2505–2512 (2008).
142. Theriot, C. M. *et al.* Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to *Clostridium difficile* infection. *Nature Commun.* **5,** 3114 (2014).
143. Weingarden, A. R. *et al.* Microbiota transplantation restores normal fecal bile acid composition in recurrent *Clostridium difficile* infection. *Am. J. Physiol. Gastrointest. Liver Physiol.* **306,** G310–G319 (2014).
144. Bernstein, H., Bernstein, C., Payne, C. M., Dvorakova, K. & Garewal, H. Bile acids as carcinogens in human gastrointestinal cancers. *Mutat. Res.* **589,** 47–65 (2005).

Acknowledgements Work in the V.S. laboratory is supported by US National Institutes of Health (NIH) grants AI053067, AI077613, AI05135 and AI114511. Work in the A.J.B. laboratory is supported by US Department of Agriculture grant 2015-67015-22930 and NIH grants AI044170, AI096528, AI112445, AI114922 and AI117940. The contents of this Review are solely the responsibility of the authors and do not necessarily represent the official views of the NIH National Institute of Allergy and Infectious Diseases.

Author information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/28ix4vg. Correspondence should be addressed to V.S. (vanessa.sperandio@utsouthwestern.edu).

© 2016 Macmillan Publishers Limited. All rights reserved.

REVIEW

doi:10.1038/nature18850

Microbiome-wide association studies link dynamic microbial consortia to disease

Jack A. Gilbert[1], Robert A. Quinn[2–4], Justine Debelius[5], Zhenjiang Z. Xu[5], James Morton[6], Neha Garg[2,3], Janet K. Jansson[7], Pieter C. Dorrestein[2–5] & Rob Knight[4–6]

Rapid advances in DNA sequencing, metabolomics, proteomics and computational tools are dramatically increasing access to the microbiome and identification of its links with disease. In particular, time-series studies and multiple molecular perspectives are facilitating microbiome-wide association studies, which are analogous to genome-wide association studies. Early findings point to actionable outcomes of microbiome-wide association studies, although their clinical application has yet to be approved. An appreciation of the complexity of interactions among the microbiome and the host's diet, chemistry and health, as well as determining the frequency of observations that are needed to capture and integrate this dynamic interface, is paramount for developing precision diagnostics and therapies that are based on the microbiome.

The role of individual species of microbes in infectious disease has been known since the work of microbiologists Robert Koch and Louis Pasteur in the nineteenth century. Yet the part played by complex communities of microbes (known as microbiotas) in providing fertile ground for infections and in setting the stage for non-communicable diseases has been appreciated only in the past decade. The gut microbiota, for example, has been linked to a variety of conditions, some of which are predictable (irritable bowel syndrome[1] and inflammatory bowel disease (IBD) in adults[2] and children[3]), whereas others are intriguing (obesity[4,5], cardiovascular disease[6], colon cancer[7] and rheumatoid arthritis[8]) or truly surprising (major depression[9], Parkinson's disease[10] and autism spectrum disorder[11]).

Many ways in which the microbiota might drive disease have been identified, but their relative importance is yet to be determined. For instance, the taxonomic composition of the microbiota might be most important, and this could be influenced by the overall diversity of species or by the presence of particular taxa, either of which can distinguish healthy individuals from those with disease states. If the collective genes of the microbiota (the microbiome) are more important, the overall genetic diversity or genetic composition, or even specific genetic lineages or metabolic pathways, might play a crucial part in shaping a disease. How such genes are expressed as transcripts and proteins could also have an effect. If the metabolome — the set of chemicals produced by the microbiota and host — is of overriding concern, whether different communities of microbes could lead to the same metabolic and immunological consequences should be considered. Overall, the molecular states of the microbiome probably interact through myriad feedback mechanisms that constantly respond and react to one another to produce the observed disease outcomes.

This Review describes the ways in which the microbiota and the microbiome, as well as specific functions of both, have been linked to various diseases. It also looks at some of the technical and conceptual pitfalls that must be avoided when designing studies that investigate these links. Such issues become compounded when studies are scaled up to cover tens of thousands of people over time and when they are designed to understand subtle and systems-level effects that result from the interactions of many factors. Microbiome-wide association studies (MWAS)[12], which capture this scale and the multidimensional interactions, and provide a means of capturing complex interactions to predict practicable links between microbial systems and disease states. MWAS can link whole microbiomes or their features to phenotypes such as disease, with appropriate controls for composition of the microbiota and unusual statistical characteristics of microbiome data sets. Although MWAS are somewhat analogous to genome-wide association studies (GWAS), the microbiome contains many more genes than does the host genome, and its composition changes over time within a person (Box 1). MWAS are useful for untangling the mechanisms that link communities of microbes and their functions to disease, although most clinical applications are yet to be fully realized. To achieve this, model systems should be devised and implemented that allow the testing of hypotheses on isolated and combinatorial functions of microbes and interventions for capturing mechanisms of action. Such systems should also enable these ideas to be applied more generally to the complex communities of microbes that inhabit the body.

Microbial biomarkers

The human microbiota is the collection of microscopic organisms that live in the body, and it contains representatives from all domains of life: the archaea, the bacteria and the eukarya. Viruses, including bacteriophages, are not always encompassed by the definition of the microbiota. They probably should be, however, because they can shape the structure of the community through top-down ecological control and they have their own effects on the immune system of the host[13]. Most approaches to identifying microbial changes have applied biomarker discovery to test for differences between people with the conditions of interest and controls. Changes in the structure of the microbiota that are associated with disease states can occur at any taxonomic rank and along any relevant branch of the phylogenetic tree. For example, changes at the phylum level have been reported in human obesity[5] and IBD[2], and strain-level associations have been made with the metabolism of drugs in humans[14]. For instance, the risk of colon cancer in mice increases in the presence of particular strains of *Escherichia coli* that express a gene cluster that produces a genotoxic secondary metabolite called colibactin[15]. Between these extremes, changes

[1]Department of Surgery, University of Chicago, Chicago, Illinois 60637, USA. [2]Department of Pharmacology, University of California San Diego, La Jolla, California 92093, USA. [3]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, California 92093, USA. [4]Center for Microbiome Innovation, Jacobs School of Engineering, University of California, San Diego, La Jolla, California 92093, USA. [5]Department of Pediatrics, University of California, San Diego School of Medicine, La Jolla, California 92093, USA. [6]Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, California 92093, USA. [7]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington 99354, USA.

© 2016 Macmillan Publishers Limited. All rights reserved.

BOX 1

Principles of microbiome-wide association studies

Microbiome-wide association studies (MWAS) are similar in concept to GWAS: the goal of both is to link a complex collection of features (for example, species or genes) to phenotype. However, there are important differences between the two. First, there are many more microbial genes than human ones, with some studies estimating that there are more than 100 microbial genes for every human gene[24,111–113]. Consequently, the issue of multiple comparisons is of greater importance to MWAS. Second, all individuals share almost the same collection of human genes but their dissimilarity in microbial species and microbial genes is much greater[24,112]. Third, genes in the human genome can be counted easily but most microbiome data comes in the form of relative abundance. Compositional statistics therefore apply and the data cannot be represented in familiar Euclidean spaces. As a result, microbiome analyses are very prone to misinterpretation. For instance, it is impossible to infer the growth or decay of microbes purely on the basis of relative abundance data because the growth of one species could also be explained by the decline of all other species. Last, whereas the human genome is essentially fixed within an individual (except in special cases such as the immune system and cancer), the microbiome of each person changes profoundly throughout his or her lifetime. Several designs for MWAS link the overall microbiome to specific phenotypes. A number of important questions must therefore be asked when designing MWAS.

- **At what level will the microbiome or microbiota be assessed?** MWAS can be carried out using species, genes, functional categories of genes or, less frequently, transcripts and proteins as features. Metabolome-wide association studies are also possible, and they can be carried out at the level of individual spectra, groups of related spectra or pathways. These analyses often give different results; for example, in the Human Microbiome Project, pathway-level analysis of the shotgun metagenomic data suggested that much less variability existed between people than did taxon-level analysis.
- **Will the microbiome be examined in terms of overall variation or as a collection of individual features?** Techniques for reducing the dimensionality of the microbiome include: clustering, principal coordinates analysis (PCoA) with a variety of distance metrics, principal component analysis, correspondence analysis, factor analysis and discriminant analysis. In clustering analyses, which include enterotyping, samples are grouped into clusters. The resulting clusters are then tested for association with a phenotype (for example, whether the resting levels of blood glucose are identical in each cluster). During dimensionality reduction, one or more axes are discovered through a supervised or an unsupervised approach, and the dependence of phenotype on locations along these axes is tested, for example, by correlation approaches. Supervised approaches such as discriminant analysis make use of phenotype labels and provide the projection of the data that best separates these class labels. Statistical tests of location on the resulting axes must therefore be used with caution because even small departures from the random model can lead to apparent separation when there is none. Unsupervised approaches such as PCoA use only the intrinsic similarities and differences in the samples; however, they may not reveal separation by phenotypic state even when it exists (because it could come only in later principal axes).

  Techniques for associating individual features of the microbiome with phenotype, including appropriate statistics for repeated measures, are Metastats[114], DESeq2 (ref. 115) and ANCOM (analysis of composition of microbiomes)[116], as well as various machine-learning approaches such as Random Forests[117]. Unfortunately, it is also challenging to infer differentially significant species in compositional data sets. Many state-of-the-art tools make assumptions about the underlying data to identify significantly different species. Analysts need to gauge the assumptions given by each tool before applying them to their data sets because these assumptions are typically not true of real-world data.
- **What corrections will be performed for multiple statistical comparisons, sparsity and compositionality of the data, and other features of microbiome and related data sets?** Often, associations will be sought between the microbiome (as a whole or as a collection of features) and measures of phenotype. In many of these studies, the differences between phenotypes can be described by a select few features. Conventional statistical tests can be confounded by the underlying ecology. For instance, multiple microbes can share the same functional roles. As a result, differences in microbial abundances might yield the same phenotype. Analyses should be separated into planned analyses (those chosen before the analysis) and ad hoc analyses (those performed after); ad hoc analyses should be considered to be exploratory rather than formal statistical tests.
- **How will causality be established?** Causality can be approached in a number of ways: through prospective longitudinal studies that demonstrate that a microbial or metabolic change precedes the disease phenotype; the demonstration that a clinical manipulation of the microbiome affects the disease process; preclinical work in mice or other animal models that demonstrates the plausibility of a mechanism; or establishment of the activity of chemical products of the microbiome that are linked to specific microbes or the genes that produce them. Studies that combine animal models with proof-of-relevance in people are especially effective, although they are rare.

at the genus level are useful for many applications, including microbial source tracking[16] and, more controversially, defining enterotypes, which are classifications of types of microbial communities in the gut[17].

Taxonomic biomarkers

Most studies have focused on identifying single organisms as biomarkers, but separating collections of samples on the basis of similarity between communities has also been useful for a wide range of diseases, including IBD[18]. However, the extent to which the choice of metric for pairwise comparisons of communities can influence the result is not widely appreciated. The fit between the data and a statistical model is often used to assess the validity of the technique. But when a collection of samples is highly heterogeneous, which includes situations as simple as the collection of skin samples from different people, models that better fit the data in the original data set might provide no clear biological interpretation. Importantly, this problem cannot be overcome by collecting more data because using the incorrect statistical model can obscure results that can be clearly determined, even with limited numbers of DNA or RNA sequences[19]. The choice of distance metric, level of taxonomic resolution or a particular taxon to focus on can involve dozens of further, implicit, comparisons that also must be accounted for statistically.

The identification of interactions between microbes is essential for microbial ecology. Correlation networks have proved useful for distilling relevant links from a morass of potential interactions. However, interpretation is still complex for two reasons. First, the abundance of specific microorganisms in each microbiota is sampled through a multinomial distribution, which leads to large numbers of negative correlations and induces a substantial bias in network topology. Second, taxonomic data are extremely sparse: most samples have zero abundance of a particular organism. Because of these correlation problems, network analyses can

© 2016 Macmillan Publishers Limited. All rights reserved.

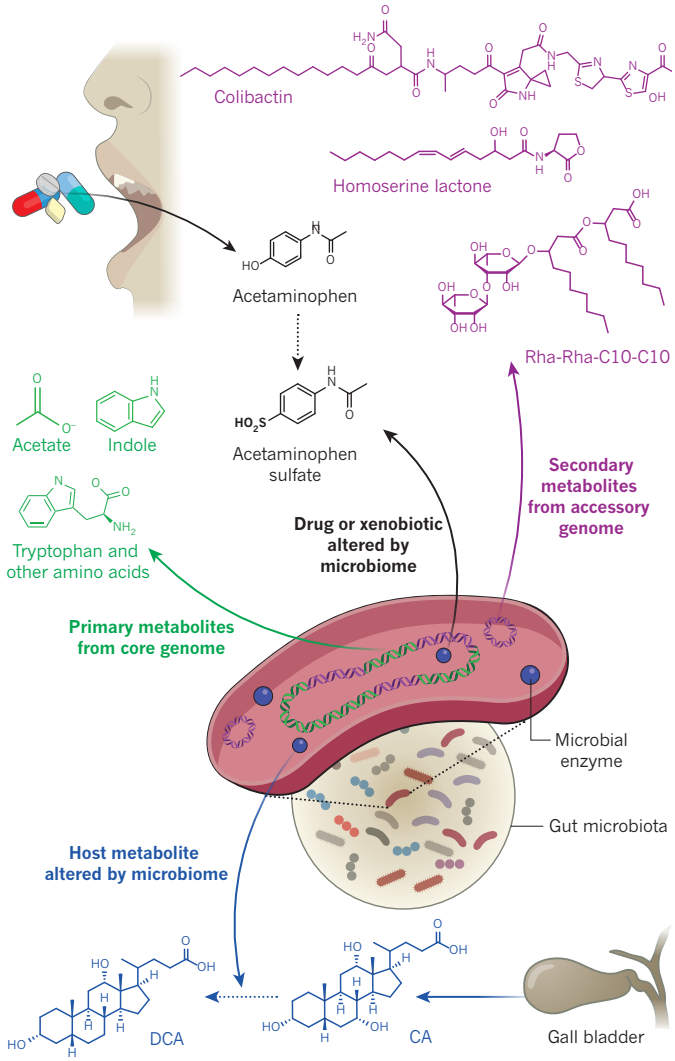


Figure 1 | Sources of metabolites from the human microbiome. The core physiology of the microbial cells that make up the microbiome can produce by-products and intermediates that affect health, including short-chain fatty acids (such as acetate) and tryptophan metabolites. Secondary (or specialized) metabolites are produced from accessory genetic elements that are often transferred horizontally between microbes. Some of these metabolites, including colibactin[15] and rhamnolipids[109] (Rha-Rha-C10-C10), are known to cause disease. Microbes can also alter metabolites that are produced by the host, such as bile acids[110] (CA, cholic acid) and even drugs that are consumed, such as acetaminophen (paracetamol)[61]. DCA, deoxycholic acid; Rha, rhamnose.

be inherently flawed[20,21]. Despite such limitations, taxonomic correlation networks have identified microbial interactions that are linked to disease, including beneficial and harmful networks of microbes that are associated with Crohn's disease[18].

These examples of successful biomarker discovery have yet to provide standard guidelines; however, they have produced interesting findings. For example, a higher level of taxonomic resolution is not always better. 16S ribosomal RNA operational taxonomic units (OTUs), which are clusters of sequences that are defined by sequence identity, at the species level are best for matching samples, yet this taxonomic level actually decreases the accuracy of classifying individuals as lean or obese[22]. The level of resolution is therefore dependent on the context.

Functional biomarkers

Shotgun metagenomics, the sequencing of fragments from total DNA rather than of specific genes, provides more-complete information about the microbial community and enables many powerful analyses, although the choice can be bewildering, even to experienced researchers in the field. As well as identifying taxa down to the level of strains or genomic single-nucleotide polymorphisms (SNPs), DNA sequences can be grouped into many functional classifications using databases such as KEGG (Kyoto Encyclopedia of Genes and Genomes), COG (Clusters of Orthologous Groups of Proteins), GO (Gene Ontology) and EggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups). Metagenomics studies[23,24] commonly show a surprising consistency in functional profiles, although the limited variation that does exist can often be explained by taxonomy. Studies that separate samples of interest from controls at different functional resolutions, are yet to be adequately performed, however. Shotgun metagenomics seems to outperform amplicon-based taxonomic analysis in the identification of individuals (compare ref. 25 with ref. 26). Re-analysis of 16S rRNA amplicon data using oligotyping[27], a technique that is based on the fine detail of polymorphisms, has improved resolution, and this is demonstrated by its ability to identify sexual partners through shared sequences. No examples are thought to exist in which shotgun metagenomics has been able to identify a medically relevant trait that could not have been revealed through taxonomic analysis alone, although the potential for doing so is high.

Integrating human metagenomic and metabolomic profiles has great potential for discriminating between disease traits (Fig. 1). The ability to systematically link the variance in metabolomic data between samples with changes in the composition and structure of communities of microbes from the same samples enables not only improved resolution but also the potential to infer the mechanisms that produce observed trends[28]. This potential is highlighted by a study[29] that shows how the microbiome alters bile-acid metabolite profiles during the establishment of *Clostridium difficile* in mice. Similarly, the ability to link metabolite profiles in urine and blood serum to microbial metabolism in the gut can help to synthesize links between dysbiosis (an imbalance of microbes in the body) and the onset of neurological symptoms that are associated with conditions such as autism spectrum disorder in a mouse model[30]. Metaproteomics is also enabling the identification of new biomarkers. Proteins such as L-lactate dehydrogenase and arginine deiminase, as well as those that are involved in the synthesis of exopolysaccharides, iron metabolism and the immune response, seem to be indicative of a healthy human oral cavity[31]. The combination of microbial community profiling with metabolomics and proteomics has precipitated understanding of how the microbiota responds to specific disease states, including IBD[32–34]. The combined findings reveal specific species (for example, *Faecalibacterium prausnitzii*), proteins and metabolites that are involved in the metabolism of butyrate and bile acids, which can be used to differentiate between individuals with inflammation of the ileum that is the result of Crohn's disease and those with inflammation of the colon and a healthy gut. In another example[35], children with non-alcoholic fatty liver disease show a significant increase in Gammaproteobacteria and *Prevotella* as well as in levels of ethanol and certain short-chain fatty acids (SCFAs), which leads to an increase in energy production and a decrease in the metabolism of carbohydrates and amino acids and in the activity of the urea cycle and urea transport systems.

From correlation to causation

A crucial challenge for the field is to move beyond associations between the microbiome and specific clinical states towards the establishment of causality. The importance of MWAS with large cohorts in determining causality should not be underestimated. The limitation of the case–control model is that it is impossible to distinguish whether the microbiome drives the disease, the disease drives the microbiome or if both are modified by a confounding factor. For example, a lack of replication of the microbiota differences that separate people with type 2 diabetes from controls in Chinese and European cohorts was found to be due to variation in the levels of usage of the drug metformin, which is used only in the disease state and with different frequencies in the two populations and which had a large and unanticipated effect on their microbiotas[36]. Consequently, the effect that had been attributed to the disease was

© 2016 Macmillan Publishers Limited. All rights reserved.

actually the result of the treatment.

Several popular methods exist for identifying causality, each of which has specific strengths and weaknesses. Prospective longitudinal studies, such as of the CHILD (Canadian Healthy Infant Longitudinal Development) birth cohort, allow researchers to test whether changes in the microbiome precede or follow the development of disease. Such studies are expensive, however, and can require large populations to capture rare events. If it is difficult to continue to collect samples, the study population can also be affected by attrition. Intervention studies, in which a deliberate clinical event such as the administration of a drug is used to drive change in the microbiome and phenotypes, are useful, but it is often unethical to withhold treatment from a control group to isolate the effect of the specific intervention. Interventions such as faecal microbiota transplantation also face substantial regulatory hurdles, especially in the United States. The comparison of identical and non-identical twins can be valuable for unravelling genetic differences in the host: causality can be established because the microbiome is not known to modify the inheritable host genome. However, such cohorts are difficult to assemble and privacy issues can be considerable, especially when the same twins are used in many studies. Animal models can be helpful for establishing mechanisms, but the quantitative importance of these mechanisms for human disease is often less clear. For example, the demonstration that faecal microbiota transplantation from people who are lean or obese to germ-free mice can confer differences in adiposity indicates that microbes can affect this phenotype, but it does not establish that transplantation can affect the weight of obese people[37].

The metabolome reveals important microbial activities

Metabolomic biomarkers are especially useful for diagnostics because changes in metabolism can be rapid and can reveal the physiological state of both the host and its microbiota. Such biomarkers are also the end products of the metabolism of microbes and they can provide mechanistic explanations for particular associations between microbes and disease. The metabolome is being characterized through metabolomics (the study of the complete repertoire of molecules in the body, which is analogous to genomics, the study of the complete repertoire of genes in the genome), metabonomics (the comparison of general metabolomics profiles with their many unidentified compounds, rather than the comparison of specific metabolites within profiles) and exposomics (the study of cumulative exposures to molecules from the environment)[38–40]. A crucial challenge for the characterization of these molecules is that only about 1.8% of the chemical data that can be collected with mass spectrometry can be annotated[39,41]. Unlike the genomics community, the mass-spectrometry community lacks adequate mechanisms of knowledge dissemination that enable data reuse. To overcome this challenge, the community is developing a plethora of resources to store data from mass spectrometry, including databases such as MassBank, METLIN, MetaboLights and the Human Metabolome Database (HMDB), the Metabolomics Workbench platform and the software OpenMS. Its efforts have also led to GNPS (Global Natural Products Social Molecular Networking), the first crowdsourced platform that enables the community-driven curation of mass-spectrometry data and dissemination of existing knowledge of mass spectrometry in the public domain[40,41]. Ultimately, these databases and infrastructures for analysis will allow the estimation of metabolite flux from the genomes to enable prediction of the overall function of communities of microbes[42]. Although a few strains of gut microbes are pathogenic, most are harmless or beneficial to health; similarly, some molecules that are produced by microbes are detrimental to the health of the host[43], but most are innocuous or even beneficial[44]. Metabolites are particularly important agents of the human microbiome. This is because molecules that are produced by the microbiota can cross epithelial barriers more freely than the microbes to cause systemic effects at distant sites in the body.

The small-molecule repertoire of the human microbiome consists of four groups. The first is composed of primary metabolites, which are molecules produced by the catabolic and anabolic reactions that are required for cellular growth and homeostasis. The second group comprises specialized metabolites, which includes virulence factors, secondary metabolites and natural products. These compounds are produced by accessory genetic elements that are often acquired through horizontal gene transfer, and they are designed to directly influence the cells of the host and other microbes (Fig. 1). Knowledge of changes in the secondary metabolites can be useful for understanding toxins, quorum sensing and beneficial secondary metabolites of food such as lycopenoids and carotenoids. The third group is composed of metabolites produced by cells of the host or from exogenous sources that are directly modified by microbial enzymes to create unique chemical products. Knowledge of changes in this group can be useful for understanding how microbes modify products of the host's metabolism. The final group is the exposome, which describes the chemistry and metabolites that are encountered through exposure to personal-care products, medical intervention, food or the environment. Knowledge of changes in this group is especially useful for understanding how compounds that are applied to the body, whether intentionally or unintentionally, can trigger toxic responses or can be modified into forms that differ in activity from the originally applied compound. Although decades of research on primary metabolism have led to a good understanding of these four groups of metabolites, the specialized metabolome of microbes is a veritable sea of unknown chemistry[45,46].

Linking metabolomes to health and disease

Evidence is accumulating that the metabolic output of the microbial metabolome has a direct impact on human health. Significant opportunities exist to elucidate the mechanisms that result in this effect. However, current methods of chemical annotation can identify only a small fraction of detected metabolites within the metabolome[41], and models for testing hypotheses about the interactions among microbes, their molecules and the host are challenging to use.

The best-known examples of microbiome-derived primary metabolites that affect human health are probably the SCFAs. SCFAs such as acetate, propionate and butyrate are produced through the fermentation of dietary fibre by gut microbes and then absorbed by epithelial cells, which provides them with energy[44,47]. Defects in the production of SCFAs have been linked to many conditions, including IBD[48,49], although it is unclear whether testing for SCFA levels *per se* has clinical value.

The development of germ-free animal models has been very useful for identifying primary metabolites produced or altered by the microbiota of the host. A comparison of metabolomes from germ-free and colonized mice revealed that indole-3-propionic acid and other products of tryptophan metabolism are found only in mice with an intact microbiota and are associated with the presence of *Clostridium sporogenes*[50]. These tryptophan metabolites are thought to affect neuronal signalling in the gut and brain[51]. But their role in human health remains elusive.

Some specialized metabolites from the human microbiota are known to cause disease. For instance, colibactin induces double-strand breaks in the DNA of human cells[52]. The genetic machinery for the production of colibactin can be transferred from pathogenic to non-pathogenic strains of *E. coli*. Colibactin is associated with colorectal cancer in mouse models[15] and provides an example of how commensal microbes of the gut can harbour or acquire specialized metabolites that can result in disease[53]. The true pathogenic elements within the human microbiota might be the genetic islands encoding specialized metabolites that circulate within the microbial ecosystem, rather than the core genomes of pathogenic species (Fig. 1). The prevalence of these genetic islands could be associated with the prevalence of microbiome-associated diseases; it is here that the interface between GWAS and MWAS can best be understood. Tests at the level of single genes, such as for genes that are necessary for colibactin production, might prove more useful for identifying preventive treatments than would tests for the presence or absence of specific taxa, akin to the way that levels of glucose or insulin are measured for diabetes.

Integration of multi-omics studies

To comprehensively understand the role of the human microbiome and

© 2016 Macmillan Publishers Limited. All rights reserved.

its metabolome in health and disease, integrative analyses are needed that apply 'omics' techniques to animal or other empirical models. Integrative analysis can help to identify the effect of treatment with antibiotics on the gut microbiota during infection with *Clostridium difficile* in both mice and people[54]. A multidisciplinary approach that employs mathematical modelling, 16S rRNA gene sequencing, metagenome sequencing and animal models identified how microbiotas can help the hosts to resist *C. difficile* infection, which led to the identification of *Clostridium scindens* as a candidate for resistance to infection in mouse models. In a study of 24 people who took antibiotics while undergoing chemotherapy[54], half had active *C. difficile* infections, which suggests that there is an association between *C. scindens* and resistance to infection with *C. difficile*. Prevention of *C. difficile* infections through transfer of *C. scindens* to animals that were undergoing treatment with antibiotics confirmed this role[55]. Metagenomic and metabolomic-based findings[56] have been used to identify the importance of bile acids in this resistance to infection, and subsequent experiments showed that certain levels of specific bile acids were associated with resistance to *C. difficile* during treatment with antibiotics. This work is an excellent example of how a comprehensive approach to microbiome analysis can link the microbiome to disease. The next step is to translate such findings into clinically useful tests.

The resident microbiota of the human gut has an important role in modulating the efficacy and toxicity of pharmaceuticals[14,57]. Variability in the microbiomes of individuals[58] leads to differences in the metabolism of drugs and therefore in effective dose availability and side effects. Simultaneous measurement of variability in both the microbiome and the metabolome will play an important part in identifying causative mechanisms of xenobiotic metabolism. The role of microbiome-associated drug toxicity is exemplified in the treatment of colon cancer with irinotecan, which resulted in decreased efficacy of the drug in 40% of treated individuals[59]. Irinotecan is reactivated in the gut by microbial β-glucuronidase enzymes, which leads to diarrhoea and prevents administration of the appropriate dose. Inhibitors that modulate the activity of the commensal microbiota by specifically inhibiting the β-glucuronidase enzyme in bacteria are in clinical trials[59]. This represents a precedent for the translation of metabolic mechanisms of the human microbiota into clinical applications and highlights the importance of investigations in the fields of pharmacomicrobiomics[60] and pharmacometabonomics[61,62]. In conjunction with testing for the genes that encode these enzymes, inhibitor-based therapies could increase the efficacy of irinotecan, although the diagnostic and therapeutic system this approach requires has yet to be demonstrated. Advanced data acquisition and computing and streamlined analysis pipelines are enabling multi-omics analysis to be performed on clinically relevant timescales, and the adaption of multi-omics microbiome analysis in the clinic will probably emerge within the next decade[63]. Future prospects should also reflect on how inhibiting specific enzymes of commensal microbes affects the overall activity and structure of the gut microbiota in the longer term.

Dynamics of the microbiome

Although many MWAS take a case–control approach, understanding how the microbiome as a whole changes remains a challenge. Relatively few studies have assessed the whole microbiome at many time points; such studies point towards using dynamic — rather than static — features as the input for MWAS. It is challenging to capture the dynamics of an invisible microbial world through snapshots of its current state. However, the situation has vastly improved in the past 15 years, during which DNA sequencing costs dropped by a factor of about one million. By increasing the frequency and depth of observations, the rate and directionality of the transfer of bacteria between ecosystems is starting to be inferred.

Assessing the transfer of microbes between environments

The application of microbial survey techniques to built environments and the people who inhabit those spaces has shown the utility of high spatiotemporal resolution for inferring interactions between people and surfaces in the environment at the microbiome level[64]. But even with daily sampling and observations at multiple sites on each individual (such as the nose, the hand and the foot), as well as their pets and surfaces in their home, it is still difficult to make more than comparative statements about the microbial similarity of surfaces and changes in this similarity over time. Higher-resolution temporal analysis, such as hourly sampling[65], can improve appreciation of the successional dynamics of these communities. These tools have not yet been applied to understanding how consistently specific components of the microbiota are transferred to or from people, let alone within the body. Alternative approaches, such as using differential coverage of parts of the microbial genome to infer activity in a single sample[66], have great promise for directly revealing activity, but samples still need to be assessed across time points because the activity of microbes can change rapidly in response to conditions. Direct monitoring of the transfer of microbes between environments and of the rapid dynamics in those environments will require a substantial improvement in the determination of genotypic resolution and temporal and spatial sampling. Near real-time microbial epidemiology is being demonstrated with genotypic resolution (at the strain level) through the rapid genomic sequencing of individual species of pathogenic microbes in hospital settings[67]. It is essential that this technology is developed to be more applicable to entire communities of microbes, especially because the most important inputs to MWAS might not be the relative abundance of each microbe or gene at a single time point but rather the variations in particular species over time, as well as their co-variations in linked environments.

Tracking pathogenic infections

Clinical application of MWAS inspires a vision of a future in which the studies are used to track entire communities of microorganisms involved in the complex 'pathobiomes' that are associated with different disease states. For example, the transfer of bacteria from mother to child might be tracked and augmented by personalized microbial therapies that range from vaginal innoculation[68] to customized prebiotic and probiotic supplements that are based on breast milk[69]. This would require automated approaches to quantify the abundance and composition (at serotype resolution) of whole communities of bacteria, as well as rapid deployment of MWAS techniques to determine current health status or to predict future health status from the trajectories. Although such sensors are not yet available, key platforms are being developed that will provide a substantial improvement on existing systems[70]. However, real-time interpretation of the vast quantity of data that are produced by these sensors will require a radical improvement in automated data processing. This will demand the integration of statistical modelling, high-performance computing and engineering to enable high-throughput transfer, interpretation and visualization of spatiotemporal data. Despite the limitations of existing correlation network techniques (that is, their sparsity and compositionality), network analysis has helped to uncover real associations in complex data. One useful example of the prediction of interactions, and subsequent validation of the prediction through empirical observation, comes from marine microbial ecology: network analysis has been applied to microbial-community sequence data to predict an interaction between an acoel flatworm (*Symsagittifera* sp.) and a green microalga (*Tetraselmis* sp.), and the finding was subsequently validated using microscopy[71]. An example in people is the use of correlation network analysis to demonstrate the connectivity of organisms in the microbiota of human milk. Cooperative and opportunistic subgroups have been identified in which the opportunistic pathogenic species could, in principle, be suppressed through competitive exclusion[72], pointing to therapeutic approaches based on probiotics (that directly introduce beneficial competing microorganisms) or prebiotics (that encourage the growth of beneficial microorganisms). The movement of microbes between environments cannot be captured by these methods, but the ability of these microorganisms to establish themselves and proliferate on arrival can be inferred through an understanding of the ecological network of their destination and their ability to incorporate. In early life, the shifting sands of an infant's microbiota can lead to an increase or decrease in the colonization success of particular microorganisms[68]. These dynamics have been tracked using longitudinal

© 2016 Macmillan Publishers Limited. All rights reserved.

characterization in work that has demonstrated the correlations between the microbiota of mother and child, especially after vaginal delivery, as well as the influence of this interaction on the transitional succession of microbial ecology in the child's gut[73]. The application of longitudinal design to MWAS would significantly improve the ability to understand the complex linkage between the microbiome and disease, and it would also improve knowledge of the link between environmental exposure and health outcomes through MWAS-enabled epidemiological investigations. Projects such as the Integrative Human Microbiome Project (iHMP)[74] are beginning to apply these approaches to larger populations.

The visualization of complex longitudinal data

Visualization improves the interpretation of data and could help to guide clinical decision-making[75]. For example, temporal dynamics could be observed in the human gut following a faecal microbiota transplant to treat a *C. difficile* infection[76], and the successional dynamics of infant microbial development could be explored[77,78]. Better visualization can also help to define the stability, resistance to perturbation and resilience to change of microbial communities; however, the quality of the initial experimental design is important. Healthy adults have unique microbial dynamics, yet patterns of stability and resistance show elements of similarity, which hints at the potential for universal ecological rules that define these relationships between individuals[79]. Determining the frequency at which longitudinal samples should be taken to capture the dynamics that are relevant to a specific disease state is an open problem[80]. For instance, two studies with different sampling intervals[81,82] found conflicting results with regards to the stability of the microbiota during pregnancy, although differences in dietary intervention could have confounded the patterns. Capturing the temporal dynamics of specific characteristics, such as the level of glucose in the blood or behavioural traits, also presents a constant challenge. The frequency at which various types of data show patterns that enable the integration and mechanistic prediction of microbial interactions should be considered. In an era of precision medicine, an understanding of when and how often different sources of information must be acquired to enable the appropriate integration of data is paramount. At best, inappropriate sampling frequencies fail to produce correlations even when mechanistic interactions exist; at worst, they produce misleading information, which might lead to the identification of incorrect biomarkers or therapeutic targets.

From explanation to prediction

The microbiome, or even the microbiota, could be used to predict the onset of disease before it occurs and to guide individualized therapies.

Stratification on the basis of the microbiome

The stratification of people for treatment holds considerable promise. For example, variation in the toxicity of acetaminophen (paracetamol) in the liver is largely caused by differences in how the drug, which is an analogue of the naturally occurring amino acid tyrosine, is metabolized through the tyrosine sulfonation pathway[61]. Similarly, the effectiveness of digoxin depends on whether the gut of an individual contains specific strains of *Eggerthella lenta*, the plasmids of which encode an enzyme that rapidly degrades digoxin and renders it ineffective[14]. Similar stories are emerging for many other classes of drug, which suggests that incorporating the gut microbiome into the stratification of participants in clinical trials and the prescription of medication could be of great value. An especially interesting example is the emerging relationship between trimethylamine *N*-oxide (TMAO) and cardiovascular disease. People can metabolize choline, which is found in dietary sources such as red meat and cheese, in a variety of ways. One such pathway is catalysed by groups of bacteria that are found only in some individuals: choline is metabolized to trimethylamine, which is then oxidized to TMAO, a compound that contributes to the formation of atherosclerotic plaques through mechanisms not yet well understood[83], although work in mice suggests a possible pathway[6]. Inhibition of the enzyme that produces TMAO or targeting relevant bacteria could therefore provide potent weapons against heart disease.

BOX 2

Integrating the host genome into MWAS

Although, intuitively, the host genome was thought to be important in shaping the microbiome, evidence to support this had been lacking. Single genes have been known to exert large effects on the gut microbiome in mice; for example, the ob/ob[5,118] and Toll-like receptor 5 knockout models[119] of obesity have been well studied, and the changes in the microbiota that are induced by a single-allele mutation can even confer part of the adiposity phenotype when transmitted by oral gavage to a genetically normal mouse. Consequently, it is well established that a genetic change can trigger an aberrant microbial community that is transmissible and can transmit the phenotype. Studies of panels of mice have shown that diet has a much larger effect on the microbiota than does the host genotype[120], and this is consistent with the observation that studies consisting of only dozens of people are unable to demonstrate that monozygotic twins are more similar in composition and function of their microbiota than are dizygotic twins[23]. However, larger studies composed of hundreds of individuals are able to find a small association between host genetics and the overall microbial community[121,122]. Intriguingly, a few taxa seem to be highly heritable, notably *Christensenella*, which is associated with leanness and even leads to weight reduction when fed to germ-free mice inoculated with the gut microbiota of obese people[122].

Conversely, it might be possible to predict whether a particular diet has adverse consequences for the heart at the level of the individual rather than the population. A study[84] involving hundreds of people was able to demonstrate this potential for diabetes; it used continuous monitoring of blood-glucose levels to understand the effects of standardized meals and their dietary components. Remarkably, ice cream was less deleterious than white rice for some people's blood glucose, and differences such as these could largely be predicted by the microbiota (and not by other factors). Consequently, using the microbiota to reduce the immense variability experienced by those who receive dietary therapies holds much promise.

Several studies have substantially advanced the field towards the goal of using the microbiome or the microbiota to predict disease before it occurs. Fascinatingly, different diseases have different dynamics. Gingivitis, an inflammation of the gums that can be reversed with thorough cleaning of the teeth, shows relapse trends that are specific to individuals, which indicates that a person's unique gingivitis-causing community of microbes returns in a predictable way[85]. By contrast, many individuals carry the same community of dental-caries-causing bacteria, yet the emergence of caries can be predicted months in advance of observable clinical symptoms by monitoring changes in the microbiota[86]. Similarly, the development of rheumatoid arthritis can be predicted using both oral and gut microbial biomarkers[87]. The potential use of oral biomarkers to predict disease that emerges at less accessible sites in the body is exciting. The oral microbiota and gut microbiota share many community members, yet the structures of communities are highly distinct, and only weak associations have been found between them[24,88]. The oral cavity provides an ideal site for non-invasive sampling and biomarker testing; the ability to use the oral microbiota to predict disease, following MWAS, therefore has tremendous promise. Predictive models are also being applied to many other sites in the body and to many other conditions, including obesity[22,89], IBD[18,90] and acne[91].

An evidence scale for microbiome studies

Although many studies have reported links between the microbiome and disease, technical variation between the studies, the effects of which often

© 2016 Macmillan Publishers Limited. All rights reserved.

exceed those of the underlying biology, makes it difficult to compare and interpret MWAS findings[78]. Efforts to quantify methodological effects enable considerable progress to be made towards performing large-scale epidemiological studies of the microbiome[92,93], but the ability to determine when specific biases require studies to be analysed separately rather than together still relies on intuition.

Crohn's disease represents one of the best-studied links between the microbiome and disease. Multiple studies[18,32–34,94], including investigations of a cohort of Swedish twins ($n = 40$ pairs of twins)[32–34], have revealed the depletion of beneficial members of the microbiota (for example, *F. prausnitzii*, a producer of butyrate) in people with inflammation in the small intestine that is associated with Crohn's disease, known as ileal Crohn's disease, compared with those who have inflammation of the colon or with healthy individuals. Increases in Proteobacteria have also been seen in these and many other studies[3,18,32]. Analysis of faecal samples from the Swedish twin cohort[33] also revealed the depletion in ileal Crohn's disease of proteins required for the metabolism of butyrate, whereas metabolite analysis[34] revealed an increase in the amounts of some bile-acid metabolites and pancreatic enzymes, as well as thousands of unidentified metabolites, that could be used to differentiate people with Crohn's disease from healthy individuals.

Type 1 diabetes has been studied in many disparate but small cohorts ($n < 20$ per group per study) of newly diagnosed children[95,96]. These studies identified an elevated relative abundance of *Bacteroides* and a reduced relative abundance of *Prevotella* in those with the disease compared with controls. A longitudinal study[97] of children with a high risk of developing diabetes determined that increases in α diversity (the diversity of species at a particular site in the body) during development were slowed in children who went on to develop diabetes ($n = 4$) but not in seroconverters without clinical symptoms ($n = 7$) or in healthy children ($n = 22$). A metabolic study in a different cohort found that children who developed diabetes ($n = 50$) had lower levels of triglycerides compared with controls ($n = 67$). Seroconversion was associated with a transient increase in 2-hydroxybutryate and a decrease in ketoleucine. Some of these metabolites might have microbial origins.

Rheumatoid arthritis, a disease not typically thought of as being associated with the gut or the mouth, has been linked to the microbiomes of both. People with rheumatoid arthritis demonstrate consistently increased relative abundances of species of *Prevotella* in their oral and gut microbiotas[8,87,98]. Those with newly diagnosed ($n = 31$) and chronic ($n = 32$) rheumatoid arthritis have higher rates of periodontal disease than do healthy controls ($n = 18$), even when other risk factors such as age and smoking are taken into account[98]. Amplicon sequencing has shown that *Prevotella* and *Leptotrichia* OTUs are increased in individuals with rheumatoid arthritis, independent of their periodontal disease status[98]. Metagenomic profiling of oral and gut microbiomes has identified elevated levels of *Prevotella copri* in people with rheumatoid arthritis ($n = 115$) compared with controls ($n = 97$), as well as an enrichment in Gram-positive microorganisms, including members of the family Veillonellaceae[87]. The presence of *Lactobacillus salivarius* in the oral cavity and faeces correlates positively with antibody titres, and this microorganism was more likely to be present in active cases of rheumatoid arthritis than in controls. Treatment with disease-modifying antirheumatic drugs can partially restore characteristics of the control microbiome, including decreased levels of *Prevotella*, in individuals with rheumatoid arthritis[98].

Cardiovascular disease[99,100] has been linked to high levels of TMAO, a metabolite of phosphotidylcholine, and TMAO is strongly correlated with both atherosclerotic plaques in a mouse model[6,83] and adverse cardiovascular outcomes in people[99,100]. TMAO has been implicated in other conditions that involve the vascular system, including renal disease[101] and colon cancer[102]. Treatment with antibiotics attenuates the production of TMAO in both mice[6] and people[103] after challenge with phosphotidylcholine. Alterations have been seen in 16S rRNA amplicon sequencing profiles of adults from Sweden[99] and China[100] who have experienced cardiovascular events, although the same OTUs were not identified in both cohorts. TMAO might also modulate platelet function and the risk of developing thrombosis in people[104]. Subsequent experiments in conventional mice have confirmed that TMAO has a role in thrombosis, whereas germ-free mice seem to be protected from developing this phenotype[104]. In conventional mice, long-term exposure to dietary choline altered the composition of the microbiome, and several candidate taxa, including the families Lachnospiraceae and Mogibacteriaceae, were negatively associated with thrombosis[104]. Interestingly, the identification of the role of TMAO in cardiovascular disease began in a study of serum metabolites, and only later moved to studies of the microbiome.

The links between autism spectrum disorder and the microbiome remain controversial; although studies in people have provided statistically significant associations, they can be confounded by factors that include the diet, gastrointestinal issues and drugs. A 16S rRNA amplicon study showed that people with autism spectrum disorder ($n = 20$) had a lower α diversity than did neurotypical individuals ($n = 20$) (ref. 11). Autism spectrum disorder was associated with higher levels of *Akkermansia* and fewer species of fermenter bacteria, including *Prevotella*, *Coprococcus* and *Veillonellaceae*[11]. A study of the offspring of mice who had undergone maternal immune activation (MIA) showed that alterations occur in the microbiomes and metabolomes of such mice, including a reduction in the levels of members of the family Lachnospiraceae, which ferment SCFAs[30]. The introduction of *Bacteroides fragilis*, a common commensal microbe, led to decreased expression of 4-ethylphenylsulfate and corrected behavioural symptoms. The administration of 4-ethylphenylsulfate was sufficient to transmit symptoms of anxiety to wild-type mice[30] and led to permanent immune dysfunction[105].

Despite the emergence of some common themes such as the presence of specific taxa, overall trends in α diversity and the ability to separate cases and controls using metrics of β diversity (the differences in community composition between different samples), it is impossible to determine whether a particular condition has a smaller or larger effect on the diversity of the microbiota than another, owing to the way that individual studies are conducted. A set of standardized protocols would enable many different biological and technical effects to be placed on a scale that compares common effect sizes. The Microbiome Quality Control Project is beginning to do this for technical effects by comparing the specific effects of sample storage, DNA extraction, PCR amplification and bioinformatics pipelines, all of which can have surprisingly large effects; for example, methods and databases used in the assignment of taxonomy can have much larger effects on the apparent profile of a microbiome than does which biological specimen was examined[106]. Large-scale efforts such as the Earth Microbiome Project[107] and American Gut are beginning to address these issues by studying tens of thousands of samples using common methods. The dream would be to provide quantitative information that indicates which biological effects are larger than specific technical effects (to facilitate a rational choice for which studies to compare) and describes the directionality of effects, which would enable the use of generalized linear models to detrend for specific variables so that subtle effects can be seen against the background. For example, American Gut has observed that the age of an individual and their self-reported frequency of alcohol consumption have approximately equal statistically significant effects on the diversity of the gut microbiota: to measure the influence of one variable accurately, it is therefore necessary to detrend for the other (American Gut, unpublished observations). By contrast, body mass index (BMI) has a much smaller, although still detectable effect, on the gut microbiota, which means that controls for age and alcohol use must be applied (or the data detrended) to understand the specific effects of BMI. The development of a scale for this type of effect size would also be enormously useful for scoping out new studies: it would enable an educated guess to be made about the expected effect size of an intervention or condition from a large database of past studies of similar phenomena, and the number of participants and longitudinal sampling design (if applicable) could be scoped out rationally on the same basis, to the relief of both investigators and their institutional review boards.

© 2016 Macmillan Publishers Limited. All rights reserved.

Developing a microbial Global Positioning System

An important challenge for the field is to move beyond abstract maps of the microbiome, which enables multivariate samples to be placed in the context of other samples. It is important to understand which factors, including the host genome (Box 2), can change the microbiome from a given starting point on a 'map', as well as where the ideal endpoint would be. Such a microbial Global Positioning System (GPS) would comprise a defined start point, a defined end point and directions for how to get from one to the other and would depend on the standardization of results from microbiome studies so that each participant can be located accurately on the map and their progress tracked. It also relies on well-defined clinical cohorts that enable desirable and undesirable endpoints to be assessed. Unstratified patients have the potential to be placed anywhere on the map, and their initial location is determined, for example, by principal coordinates analysis (PCoA) of UniFrac distances between samples[108], as performed by the Human Microbiome Project and American Gut. Stratification is then performed to identify certain groups of people in different parts of the map, according to specific biomarkers, such as genes, functions, metabolites or networks of these features, and perhaps by crossing different levels of analysis. These biomarkers are then used to relocate study participants to appropriate regions of the map, which helps to suggest specific treatments that would move them from their present location to another (Fig. 2). For example, a small change in the diet might provide a subtle shift in location on the map and treatment with antibiotics might produce a larger shift whereas faecal transplantation could be considered 'teleportation'. Readout of biomarkers over time would allow the progress of each participant to be tracked from unhealthy to healthier regions of the map. Overall, many more participants would be expected to reach a healthy location on the map than would be possible with unstratified treatment, although genetic defects, intractable microbiome states or other factors might prevent the recovery of some. This vision requires a substantially faster, cheaper and more accurate readout of the microbiome across multiple levels than is possible at present, although it will provide an exceptionally powerful and clinically relevant model after it has been subjected to the appropriate regulatory processes.

Perspective

The considerable power of using the microbiome, or even the inexpensively assayed microbiota, to separate cases from controls, as well as to predict responses to treatment or the development of diseases in the absence of treatment, has already been demonstrated through carefully controlled MWAS in research settings. To further develop these techniques for robust clinical use, MWAS must be validated in larger and more diverse populations. Methodologies must also be standardized so that differences in the size of technical effects between laboratories do not outweigh differences in the size of biological effects, which can make studies difficult to combine[78,90]. This problem remains a crucial challenge to overcome and prevents findings from being developed into clinical tests.

Longitudinal studies have been especially informative in revealing microbiome dynamics that cannot be observed through a before–after model. In infants, where profound changes in the microbiota and microbiome occur in the first three years of life, a more detailed understanding of the developmental process, and deviations from it, is required to understand whether changes introduced by diet, environmental exposures, antibiotics and other factors in early life keep the microbiome on track or divert it towards danger. Similarly, moving away from taxonomic inventories towards an understanding of the genes, transcripts, proteins and metabolites of the microbiome in a multi-omics, systems-biology context is crucial for generalizing our understanding of a wide range of diseases in which the microbiome is involved, as well as for developing biomarkers that could be the basis of useful clinical tests. However, these are conflicting imperatives: multi-omics studies greatly increase the cost of analysing each sample, which means that longitudinal studies on large populations quickly become infeasible and tests are too expensive and slow to apply on clinically relevant timescales. Consequently, even higher-throughput and cheaper methods to process samples for multi-omics studies, as well as improved modelling techniques that derive systems-level dynamic parameters from fewer samples, are urgently required. These advances will rapidly bring us nearer to the dream of a microbial GPS. The Human Microbiome Project, the Earth Microbiome Project, American Gut and other large-scale efforts have already, and very effectively, provided a microbial 'map' that enables healthy and diseased samples to be placed



Figure 2 | Developing a microbial Global Positioning System to stratify individuals and to guide their treatment. An unstratified pool of individuals (black), all of whom have the same disease but with different underlying states (red, blue and grey), are stratified according to a biomarker from the microbiota, the microbiome or the metabolome (differentiated on a PCoA plot (bottom) or other analysis). This enables treatments to be chosen for each subpool, which facilitates movement from an 'unhealthy' region to a 'healthy' region of the microbial 'map'. The position of an individual in the main pool indicates the same person over time. The microbial Global Positioning System therefore enables determination of the current location of an individual in terms of their microbiome configuration, as well as a prediction of their final destination and directions for how to get there. Ideally, this moves all individuals in the pool to a healthy status (green) and microbiome, although in real-world situations no treatment will work perfectly. PC, principal coordinate.

© 2016 Macmillan Publishers Limited. All rights reserved.

in context, provided that consistent laboratory and bioinformatics methods are used. In the next few years, data that are collected using consistent protocols will enable intervention studies from many investigators to be aggregated to build a general picture of how the microbiome can change in specific directions in multivariate space. This understanding will facilitate the provision of 'turn-by-turn' directions that enable individuals to use their microbiome and perhaps even their genotype to understand where they might want to go on this map and how they can get there most effectively, in a way that preserves their lifelong health. ■

Received 7 December 2015; accepted 6 May 2016.

1. Manichanh, C. *et al.* Anal gas evacuation and colonic microbiota in patients with flatulence: effect of diet. *Gut* **63,** 401–408 (2014).
2. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104,** 13780–13785 (2007).
This study linked the microbiota to IBD and also demonstrated that various forms of the condition have distinct signatures of microbiota.
3. Lewis, J. D. *et al.* Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe* **18,** 489–500 (2015).
4. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
5. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
6. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Med.* **19,** 576–585 (2013).
7. Kostic, A. D. *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* **22,** 292–298 (2012).
This study identified high levels of *Fusobacterium nucleatum* in tissue from human tumours; the bacterium was later confirmed to cause tumours in experiments in animals.
8. Scher, J. U. *et al.* Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2,** e01202 (2013).
This paper provided the first evidence to directly link the gut microbiota to rheumatoid arthritis in people.
9. Naseribafrouei, A. *et al.* Correlation between the human fecal microbiota and depression. *Neurogastroenterol. Motil.* **26,** 1155–1162 (2014).
10. Scheperjans, F. *et al.* Gut microbiota are related to Parkinson's disease and clinical phenotype. *Mov. Disord.* **30,** 350–358 (2015).
11. Kang, D. W. *et al.* Reduced incidence of *Prevotella* and other fermenters in intestinal microflora of autistic children. *PLoS ONE* **8,** e68322 (2013).
12. Kostic, A. D., Howitt, M. R. & Garrett, W. S. Exploring host–microbiota interactions in animal models and humans. *Genes Dev.* **27,** 701–718 (2013).
13. Barr, J. J. *et al.* Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl Acad. Sci. USA* **110,** 10771–10776 (2013).
14. Haiser, H. J. *et al.* Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* **341,** 295–298 (2013).
This study provided a mechanism to underpin the high variation between individuals in efficacy of the cardiac drug digoxin, which was suspected (but not yet proven) to be linked to its metabolism by *Eggerthella lenta*.
15. Arthur, J. C. *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338,** 120–123 (2012).
16. Knights, D. *et al.* Bayesian community-wide culture-independent microbial source tracking. *Nature Methods* **8,** 761–763 (2011).
17. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
18. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15,** 382–392 (2014).
This study of treatment-naive children who had been freshly diagnosed with Crohn's disease enabled the effects of treatment to be separated from those of the condition.
19. Kuczynski, J. *et al.* Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature Methods* **7,** 813–819 (2010).
20. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8,** e1002687 (2012).
21. Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S. & Bahler, J. Proportionality: a valid alternative to correlation for relative data. *PLoS Comput. Biol.* **11,** e1004075 (2015).
22. Knights, D., Parfrey, L. W., Zaneveld, J., Lozupone, C. & Knight, R. Human-associated microbial signatures: examining their predictive value. *Cell Host Microbe* **10,** 292–296 (2011).
23. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
24. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
25. Fierer, N. *et al.* Forensic identification using skin bacterial communities. *Proc. Natl Acad. Sci. USA* **107,** 6477–6481 (2010).
26. Franzosa, E. A. *et al.* Identifying personal microbiomes using metagenomic codes. *Proc. Natl Acad. Sci. USA* **112,** E2930–E2938 (2015).
27. Eren, A. M. *et al.* Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4,** 1111–1119 (2013).
This paper demonstrated how careful analysis of exact 16S rRNA sequences that avoids clustering into OTUs can reveal fine-grained information that can be useful for forensic matching.
28. Noecker, C. *et al.* Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. *mSystems* **13,** http://dx.doi.org/10.1128/mSystems.00013-15 (2015).
29. Koenigsknecht, M. J. *et al.* Dynamics and establishment of *Clostridium difficile* infection in the murine gastrointestinal tract. *Infect. Immun.* **83,** 934–941 (2015).
30. Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155,** 1451–1463 (2013).
This study showed that the phenotype of a mouse model of autism spectrum disorder could be traced, in part, to a single molecule (4-ethylphenylsulfate) and a shift in the microbiota that can be partially restored using a probiotic.
31. Belda-Ferre, P. *et al.* The human oral metaproteome reveals potential biomarkers for caries disease. *Proteomics* **15,** 3497–3507 (2015).
32. Willing, B. P. *et al.* A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139,** 1844–1854 (2010).
33. Erickson, A. R. *et al.* Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* **7,** e49138 (2012).
34. Jansson, J. *et al.* Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* **4,** e6386 (2009).
35. Michail, S. *et al.* Altered gut microbial energy and metabolism in children with non-alcoholic fatty liver disease. *FEMS Microbiol. Ecol.* **91,** 1–9 (2015).
36. Forslund, K. *et al.* Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528,** 262–266 (2015).
This paper compared two discordant studies of microbiomes in type 2 diabetes and showed that the alleged effect of diabetes could be attributed mostly to differences in use of metformin, which has an unexpectedly large effect on the microbiome, between the two populations.
37. Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341,** 1241214 (2013).
This study demonstrated that phenotypes such as increased adiposity could be transferred from people to mice using personalized culture collections.
38. Li, H. & Jia, W. Cometabolism of microbes and host: implications for drug metabolism and drug-induced toxicity. *Clin. Pharmacol. Ther.* **94,** 574–581 (2013).
39. Nicholson, J. K., Lindon, J. C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29,** 1181–1189 (1999).
40. Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Rev. Drug Discov.* http://dx.doi.org/10.1038/nrd.2016.32 (2016).
41. da Silva, R. R., Dorrestein, P. C. & Quinn, R. A. Illuminating the dark matter in metabolomics. *Proc. Natl Acad. Sci. USA* **112,** 12549–12550 (2015).
42. Gilbert, J. A. & Henry, C. Predicting ecosystem emergent properties at multiple scales. *Environ. Microbiol. Rep.* **7,** 20–22 (2015).
43. Allen, L. *et al.* Pyocyanin production by *Pseudomonas aeruginosa* induces neutrophil apoptosis and impairs neutrophil-mediated host defenses *in vivo*. *J. Immunol.* **174,** 3643–3649 (2005).
44. Puertollano, E., Kolida, S. & Yaqoob, P. Biological significance of short-chain fatty acid metabolism by the intestinal microbiome. *Curr. Opin. Clin. Nutr. Metab. Care* **17,** 139–144 (2014).
45. Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158,** 412–421 (2014).
46. Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158,** 1402–1414 (2014).
This paper showed that the human microbiome harbours many biosynthetic gene clusters, including those required for the production of antibiotics.
47. Cummings, J. H. Fermentation in the human large intestine: evidence and implications for health. *Lancet* **1,** 1206–1209 (1983).
48. Huda-Faujan, N. *et al.* The impact of the level of the intestinal short chain fatty acids in inflammatory bowel disease patients versus healthy subjects. *Open Biochem. J.* **4,** 53–58 (2010).
49. Ríos-Covián, D. *et al.* Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* **7,** 185 (2016).
50. Wikoff, W. R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl Acad. Sci. USA* **106,** 3698–3703 (2009).
51. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science* **349,** 1254766 (2015).
52. Nougayrède, J. P. *et al. Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313,** 848–851 (2006).
53. Putze, J. *et al.* Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect. Immun.* **77,** 4696–4703 (2009).
54. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* **517,** 205–208 (2015).
55. Langille, M. G. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnol.* **31,** 814–821 (2013).
56. Allegretti, J. R. *et al.* Recurrent *Clostridium difficile* infection associates with distinct bile acid and microbiome profiles. *Aliment. Pharmacol. Ther.* **43,** 1142–1153 (2016).
57. Maurice, C. F., Haiser, H. J. & Turnbaugh, P. J. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152,** 39–50 (2013).
58. Mani, S., Boelsterli, U. A. & Redinbo, M. R. Understanding and modulating

© 2016 Macmillan Publishers Limited. All rights reserved.

mammalian-microbial communication for improved human health. *Annu. Rev. Pharmacol. Toxicol.* **54,** 559–580 (2014).
59. Wallace, B. D. *et al.* Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330,** 831–835 (2010).
This paper demonstrated that the cancer therapeutic drug irinotecan causes severe diarrhoea because of its reactivation and metabolism by bacterial β-glucuronidases; inhibiting these enzymes with a drug that targets the bacteria, rather than the host, reduces toxicity.
60. ElRakaiby, M. *et al.* Pharmacomicrobiomics: the impact of human microbiome variations on systems pharmacology and personalized therapeutics. *OMICS* **18,** 402–414 (2014).
61. Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. Pharmacometabonomic identification of a significant host–microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl Acad. Sci. USA* **106,** 14728–14733 (2009).
This study provided the first link between the toxicity of a drug (in this case, acetaminophen, a widely used analgesic) and microbial metabolism.
62. Wilson, I. D. Drugs, bugs, and personalized medicine: pharmacometabonomics enters the ring. *Proc. Natl Acad. Sci. USA* **106,** 14187–14188 (2009).
63. Quinn, R. A. *et al.* From sample to multi-omics conclusions in under 48 hours. *mSystems* http://dx.doi.org/10.1128/mSystems.00038-16 (2016).
64. Lax, S. *et al.* Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* **345,** 1048–1052 (2014).
65. Gibbons, S. M. *et al.* Ecological succession and viability of human-associated microbiota on restroom surfaces. *Appl. Environ. Microbiol.* **81,** 765–773 (2015).
66. Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349,** 1101–1106 (2015).
67. Quick, J. *et al.* Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome Biol.* **16,** 114 (2015).
68. Dominguez-Bello, M. G. *et al.* Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nature Med.* **22,** 250–253 (2016).
69. Goyal, M. S., Venkatesh, S., Milbrandt, J., Gordon, J. I. & Raichle, M. E. Feeding the brain and nurturing the mind: linking nutrition and the gut microbiota to brain development. *Proc. Natl Acad. Sci. USA* **112,** 14105–14112 (2015).
70. Biteen, J. S. *et al.* Tools for the microbiome: nano and beyond. *ACS Nano* **10,** 6–37 (2016).
71. Lima-Mendez, G. *et al.* Determinants of community structure in the global plankton interactome. *Science* **348,** 1262073 (2015).
72. Sam Ma, Z. *et al.* Network analysis suggests a potentially 'evil' alliance of opportunistic pathogens inhibited by a cooperative network in human milk bacterial communities. *Sci. Rep.* **5,** 8275 (2015).
73. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17,** 690–703 (2015); erratum **17,** 852 (2015).
74. The Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: dynamic analysis of microbiome–host omics profiles during periods of human health and disease. *Cell Host Microbe* **16,** 276–289 (2014).
75. Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A. & Knight, R. EMPeror: a tool for visualizing high-throughput microbial community data. *Gigascience* **2,** 16 (2013).
76. Weingarden, A. *et al.* Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent *Clostridium difficile* infection. *Microbiome* **3,** 10 (2015).
This paper introduced animation techniques that revealed the transformation of the whole microbiota during faecal microbiota transplantation for *C. difficile* infection.
77. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4578–4585 (2011).
78. Lozupone, C. A. *et al.* Meta-analyses of studies of the human microbiota. *Genome Res.* **23,** 1704–1714 (2013).
79. Flores, G. E. *et al.* Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **15,** 531 (2014).
80. Shade, A. *et al.* Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mBio* **5,** e01371-14 (2014).
81. DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl Acad. Sci. USA* **112,** 11060–11065 (2015).
82. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150,** 470–480 (2012).
83. Wang, Z. *et al.* Prognostic value of choline and betaine depends on intestinal microbiota-generated metabolite trimethylamine-*N*-oxide. *Eur. Heart J.* **35,** 904–910 (2014).
84. Zeevi, D. *et al.* Personalized nutrition by prediction of glycemic responses. *Cell* **163,** 1079–1094 (2015).
This study showed that individual glycaemic responses could be predicted using the microbiome; it also revealed that although population averages match conventional glycaemic-index values, the responses of individuals are highly idiosyncratic and dependent on the microbiome.
85. Teng, F. *et al.* Prediction of early childhood caries via spatial-temporal variations of oral microbiota. *Cell Host Microbe* **18,** 296–306 (2015).
86. Huang, S. *et al.* Predictive modeling of gingivitis severity and susceptibility via oral microbiota. *ISME J.* **8,** 1768–1780 (2014).
87. Zhang, X. *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature Med.* **21,** 895–905 (2015).
88. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509,** 357–360 (2014).
89. Cotillard, A. *et al.* Dietary intervention impact on gut microbial gene richness. *Nature* **500,** 585–588 (2013).
90. Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* **588,** 4223–4233 (2014).
91. Kang, D., Shi, B., Erfe, M. C., Craft, N. & Li, H. Vitamin B_{12} modulates the transcriptome of the skin microbiota in acne pathogenesis. *Sci. Transl. Med.* **7,** 293ra103 (2015).
92. Sinha, R., Abnet, C. C., White, O., Knight, R. & Huttenhower, C. The microbiome quality control project: baseline study design and future directions. *Genome Biol.* **16,** 276 (2015).
93. Sinha, R. *et al.* Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol. Biomarkers Prev.* **25,** 407–416 (2016).
94. Sokol, H. *et al. Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105,** 16731–16736 (2008).
95. de Goffau, M. C. *et al.* Fecal microbiota composition differs between children with β-cell autoimmunity and those without. *Diabetes* **62,** 1238–1244 (2013).
96. Giongo, A. *et al.* Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J.* **5,** 82–91 (2011).
97. Kostic, A. D. *et al.* The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* **17,** 260–273 (2015).
98. Scher, J. U. *et al.* Periodontal disease and the oral microbiota in new-onset rheumatoid arthritis. *Arthritis Rheum.* **64,** 3083–3094 (2012).
99. Koren, O. *et al.* Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4592–4598 (2011).
100. Yin, J. *et al.* Dysbiosis of gut microbiota with reduced trimethylamine-*N*-oxide level in patients with large-artery atherosclerotic stroke or transient ischemic attack. *J. Am. Heart Assoc.* **4,** e002699 (2015).
101. Tang, W. H. *et al.* Gut microbiota-dependent trimethylamine *N*-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circ. Res.* **116,** 448–455 (2015).
102. Xu, R., Wang, Q. & Li, L. A genome-wide systems analysis reveals strong link between colorectal cancer and trimethylamine *N*-oxide (TMAO), a gut microbial metabolite of dietary meat and fat. *BMC Genomics* **16** (suppl. 7), S4 (2015).
103. Tang, W. H. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368,** 1575–1584 (2013).
104. Zhu, W. *et al.* Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* **165,** 111–124 (2016).
105. Hsiao, E. Y., McBride, S. W., Chow, J., Mazmanian, S. K. & Patterson, P. H. Modeling an autism risk factor in mice leads to permanent immune dysregulation. *Proc. Natl Acad. Sci. USA* **109,** 12776–12781 (2012).
106. Liu, Z., DeSantis, T. Z., Andersen, G. L. & Knight, R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res.* **36,** e120 (2008).
107. Gilbert, J. A. *et al.* Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. *Stand. Genomic Sci.* **3,** 243–248 (2010).
108. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71,** 8228–8235 (2005).
109. Quinn, R. A. *et al.* Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome. *ISME J.* **10,** 1483–98 (2015).
110. Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. Bile acids and the gut microbiome. *Curr. Opin. Gastroenterol.* **30,** 332–338 (2014).
111. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312,** 1355–1359 (2006).
This study provided the first metagenomic gene catalogue of the human gut.
112. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
113. Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* **449,** 804–810 (2007).
114. White, J. R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **5,** e1000352 (2009).
115. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
116. Mandal, S. *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* **26,** 27663 (2015).
117. Knights, D., Costello, E. K. & Knight, R. Supervised classification of human microbiota. *FEMS Microbiol. Rev.* **35,** 343–359 (2011).
118. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102,** 11070–11075 (2005).
119. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328,** 228–231 (2010).
120. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab.* **17,** 141–152 (2013).
121. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).
122. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159,** 789–799 (2014).

Acknowledgements This work and the work in the authors' laboratories that it describes was supported in part by awards from the US National Institutes of Health, the US Department of Energy, the US National Science Foundation, the Alfred P. Sloan Foundation, the Crohn's and Colitis Foundation of America and the US Office of Naval Research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/28jf0c5. Correspondence should be addressed to R.K. (robknight@ucsd.edu).

© 2016 Macmillan Publishers Limited. All rights reserved.

ARTICLE

doi:10.1038/nature18609

Species-specific wiring for direction selectivity in the mammalian retina

Huayu Ding[1], Robert G. Smith[2], Alon Poleg-Polsky[1], Jeffrey S. Diamond[1] & Kevin L. Briggman[3,4]

Directionally tuned signalling in starburst amacrine cell (SAC) dendrites lies at the heart of the circuit that detects the direction of moving stimuli in the mammalian retina. The relative contributions of intrinsic cellular properties and network connectivity to SAC direction selectivity remain unclear. Here we present a detailed connectomic reconstruction of SAC circuitry in mouse retina and describe two previously unknown features of synapse distributions along SAC dendrites: input and output synapses are segregated, with inputs restricted to proximal dendrites; and the distribution of inhibitory inputs is fundamentally different from that observed in rabbit retina. An anatomically constrained SAC network model suggests that SAC–SAC wiring differences between mouse and rabbit retina underlie distinct contributions of synaptic inhibition to velocity and contrast tuning and receptive field structure. In particular, the model indicates that mouse connectivity enables SACs to encode lower linear velocities that account for smaller eye diameter, thereby conserving angular velocity tuning. These predictions are confirmed with calcium imaging of mouse SAC dendrites responding to directional stimuli.

A thorough understanding of a neuronal circuit requires a detailed anatomical wiring diagram that includes the synaptic connectivity among the component neurons. Even ostensibly subtle connectivity differences during development or between species could underlie substantial changes in circuit behaviour. This is exemplified in the direction selectivity circuit in the mammalian retina, a model neural network that engages just a few well-characterized cell types to compute salient visual information. However, the detailed synaptic connectivity among these neurons, and circuitry differences between species, has not been completely described.

Direction-selective ganglion cells (DSGCs) respond strongly to visual motion in one (preferred) direction but only weakly to motion in the opposite (null) direction[1]. Bipolar cells provide excitatory synaptic inputs to DSGCs and to densely arrayed SACs, which then inhibit DSGCs (Fig. 1a)[2,3]. SAC dendrites oriented asymmetrically to a DSGC provide feedforward inhibitory input that establishes DSGC directional tuning[4,5]. SAC dendrites are themselves directionally selective and release the neurotransmitter GABA (γ-aminobutyric acid) from synaptic terminals at their tips preferentially in response to outward (centrifugal) compared to inward (centripetal) motion relative to their soma[6–8]. Several mechanisms contribute to direction selectivity within individual SAC dendrites, but the relative importance of the mechanisms is unclear. Proposed intrinsic mechanisms include dendritic morphology[9], non-uniform chloride homeostasis[10], and active membrane conductances[6,11]. SAC direction selectivity may also rely on network interactions, such as spatially offset synaptic inputs from particular bipolar cell types[12] and reciprocal inhibition between neighbouring SACs[8,13–16].

Most anatomical analyses of SAC microcircuitry have been performed in rabbit retina. Sparse electron microscopy reconstructions in rabbit indicated that excitatory and inhibitory synaptic inputs occur along the entire length of SAC dendrites, whereas inhibitory synaptic outputs arise on the distal third[2]. We explored SAC connectivity in mouse retina using serial block-face scanning electron microscopy[17]. We discovered a previously unknown asymmetric distribution of inhibitory and excitatory input synapses onto ON and OFF mouse SAC dendrites that is fundamentally different from the connectivity in rabbit retina. We developed an anatomically constrained network model of mouse SAC connectivity that predicts new roles for synaptic inhibition in velocity and contrast tuning and receptive field structure in SACs. Finally, we confirmed these predictions by recording directionally tuned responses in mouse SAC dendrites. Our results indicate that the SAC network has adapted to meet the specific demands imposed by the mouse visual system.

Synaptic inputs are spatially offset

We annotated an ON–OFF DSGC within a conventionally stained serial block-face scanning electron microscopy volume ($50 \times 210 \times 260\,\mu m^3$) from an adult mouse retina (Extended Data Fig. 1a, b). Neurites forming conventional (inhibitory) synapses (Fig. 1b) onto this cell were back-traced to identify four SACs (2 ON, 2 OFF) located centrally in the data volume (Fig. 1c–e and Extended Data Fig. 1c–e). The morphology of each SAC was fully traced within the data volume and the locations of input and output synapses were annotated. As expected, output synapses arose along the distal third of SAC dendritic trees (Fig. 1f, g). Ribbon-type input synapses (Fig. 1b) from bipolar cells were distributed primarily along the proximal two-thirds of dendrites (Fig. 1f, g). Conventional synapses from amacrine cells (Fig. 1b) were restricted to the initial third of the dendritic trees (Fig. 1f, g). This proximal location of amacrine cell inputs differs from previous reports in rabbit retina that SACs receive reciprocal SAC inputs along their distal dendrites (Extended Data Fig. 1f)[2,8,16,18], indicating that SAC connectivity is fundamentally different in mice and rabbits. Next, we identified cells that were presynaptic to the SACs.

Bipolar and amacrine cell types presynaptic to SACs

Recent analysis of contact area shared between different OFF bipolar cell types and OFF SACs suggested a 'space-time wiring' presynaptic delay model that supports SAC direction selectivity[12]. In this model, different bipolar cell types exhibit distinct release kinetics[19–21], and

[1]Synaptic Physiology Section, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland 20892, USA. [2]Department of Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [3]Department of Biomedical Optics, Max Planck Institute for Medical Research, Heidelberg 69120, Germany. [4]Circuit Dynamics and Connectivity Unit, National Institute of Neurological Disorders and Stroke, Bethesda, Maryland 20892, USA.

© 2016 Macmillan Publishers Limited. All rights reserved

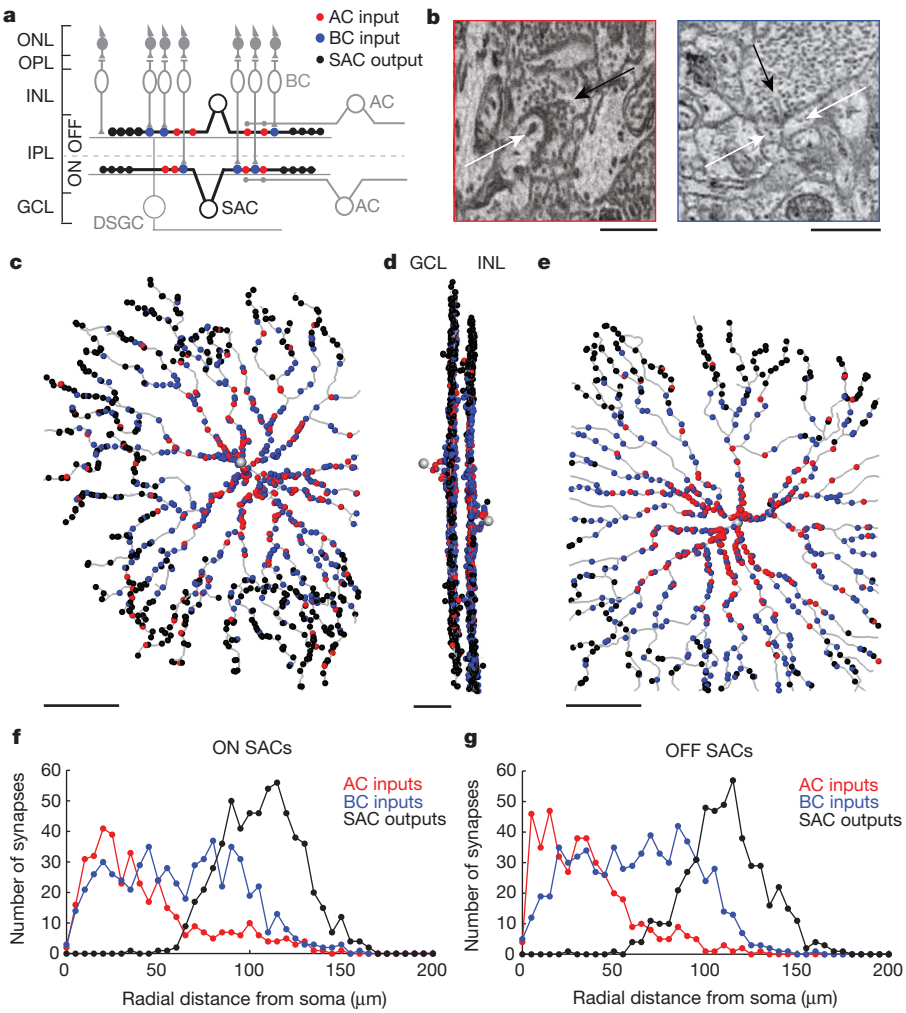


Figure 1 | Synaptic connectivity of mouse SACs. a, Schematic diagram of the direction selectivity circuitry. AC, amacrine cells, BC, bipolar cells, ONL, outer nuclear layer, OPL, outer plexiform layer, INL, inner nuclear layer, IPL, inner plexiform layer, GCL, ganglion cell layer. **b**, Representative examples of a presynaptic SAC (black arrow, left image) contacting a postsynaptic SAC (white arrow, left image) and a presynaptic bipolar cell (black arrow, right image) forming a ribbon synapse with two postsynaptic SACs (white arrows, right image). **c**, **e**, Distribution of excitatory (blue) and inhibitory (red) input synapses and output synapses (black) onto ON and OFF SACs. **d**, Horizontal view of ON and OFF SACs, whose somata reside in the GCL or INL, respectively. **f**, **g**, Histograms of radial distances from the soma for annotated synapses. Data pooled from $n = 2$ ON and $n = 2$ OFF SACs. Scale bars, 1 μm (**b**), 50 μm (**c**, **e**), 25 μm (**d**).

sustained bipolar cells (for example, BC2) provide input more proximally on SAC dendrites than do transient bipolar cells (for example, BC3a). Because our data set allowed us to positively identify synapses, we classified bipolar cell types providing input to ON and OFF SACs (Fig. 2); and we noted several differences compared to the contact-based analysis[12].

We found synapses onto the OFF SACs from all OFF bipolar cell types (BC1, BC2, BC3a, BC3b, and BC4; Fig. 2a, b and Extended Data Fig. 2) with most input from BC1, BC2, and BC3a (Fig. 2a, b). BC1 and BC3a exhibited segregated radial distributions, potentially supporting a presynaptic space-time wiring model, but BC2 overlapped with both; this overlap, regardless of BC2 response kinetics, would presumably diminish direction selectivity generated in such a model. Space–time wiring may still support direction selectivity in OFF SACs, pending characterization of type-specific OFF bipolar cell response kinetics. Our data suggest that SAC dendrites may simply sample from the available bipolar cells at a particular depth in the inner plexiform layer (IPL) (Fig. 2c), regardless of bipolar cell release characteristics.

If space–time wiring were essential for SAC direction selectivity, one would expect a similar connectivity pattern for ON SACs. Bipolar cell inputs to the ON SACs clustered into four subtypes, corresponding to BC7 and three BC5 subtypes (BC5o, BC5t, BC5i)[22] (Figs 2d, e and Extended Data Fig. 3). We found that type BC7 primarily contacted proximal dendrites, whereas BC5 inputs, collectively, were distributed more distally (Fig. 2d, e). The radial location of synapses correlated with their IPL depth (Fig. 2f). Segregated bipolar cell inputs to ON SACs could support a space–time direction selectivity mechanism, although type BC7 (which we show provide proximal inputs) exhibits transient light responses[23], counter to the model's requirements.

Next, we analysed the sources of amacrine cell synapses onto the ON and OFF SACs (Fig. 3a, b and Extended Data Fig. 4). Most inputs originated from neighbouring SACs, identified by their distinctive branching pattern and tight co-stratification with the postsynaptic SACs (Fig. 3a and Extended Data Fig. 4a, b). There was no directional preference in the absolute orientations of presynaptic SAC dendrites. Previous studies hypothesized that SAC direction selectivity could be enhanced if opposing ('anti-parallel') SAC dendrites preferentially made reciprocal connections[8,13]. To test this idea, we measured the relative angle between connected presynaptic and postsynaptic dendrites (Fig. 3c and Extended Data Fig. 5a, b). The distributions of relative angles for both the ON and OFF SACs were significantly skewed towards anti-parallel (180°) wiring (Kolmogorov–Smirnov test, $P = 2 \times 10^{-53}$; Fig. 3c). We considered whether presynaptic SAC dendrites selectively connect to opposing dendrites or whether the relative angle distribution simply reflects the inter-soma spacing between SACs. We annotated locations where the distal third of presynaptic SAC dendrites passed within 1 μm of the postsynaptic SACs and measured the relative angles between dendrites at each proximity. The proximity-based relative angle distribution was not statistically significantly different from the distribution based on actual synaptic connectivity (Extended Data Fig. 5c, Kolmogorov–Smirnov test, $P = 0.18$), indicating that the wiring arises primarily from the geometric arrangement of connected SACs. Relative angle was not correlated to the radial distance of each synapse from the respective postsynaptic SAC soma (Extended Data Fig. 5d).

Not all inhibitory inputs came from neighbouring SACs. We annotated several apparent wide-field amacrine cells that contributed synapses specifically onto the most proximal dendrites of ON and OFF SACs (Extended Data Fig. 4c, d). Wide-field amacrine cells did not

© 2016 Macmillan Publishers Limited. All rights reserved

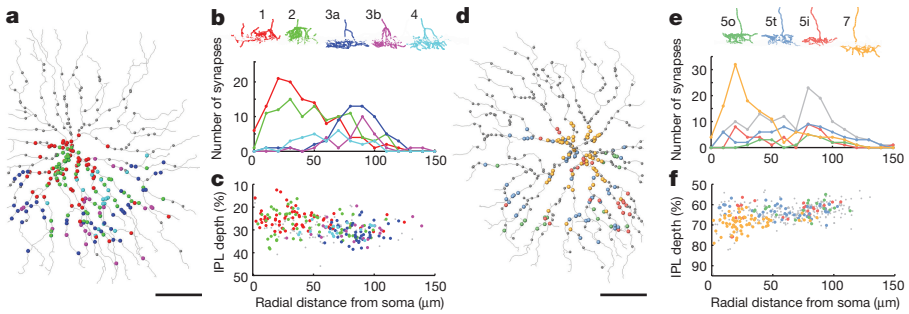


Figure 2 | Bipolar cell inputs to mouse SACs. **a**, **b**, Location of BC synapses onto an OFF SAC, colour-coded by bipolar cell type (**b**). Grey dots indicate BC synapses that were not analysed. **b**, Total OFF bipolar cell synapses ($n = 343$) onto $n = 2$ OFF SACs versus radial distance from soma. **c**, IPL depth of each synapse versus the radial distance relative to their soma. **d**–**f**, As in **a**–**c** for ON SACs, ($n = 262$) ON bipolar cell synapses onto $n = 2$ ON SACs. Grey line (**e**) indicates pooled inputs from all three type 5 bipolar cells. Scale bar, 50 μm.

co-stratify with SACs, but rather stratified close to the inner nuclear and ganglion cell layers, in contrast to a different population targeting bipolar cell axon terminals presynaptic to DSGCs[24]. We also found a few synapses from narrow-field amacrine cells, mostly onto ON SACs (Extended Data Fig. 4e)[25]. Therefore, although most proximal amacrine inputs originate from neighbouring SACs, additional inputs may selectively inhibit perisomatic compartments.

We also quantified the number and types of postsynaptic targets of ON and OFF SAC branches terminating near the centre of the data volume (Extended Data Fig. 6). We traced postsynaptic cells until they could be identified unambiguously as a ganglion cell, SAC, wide-field amacrine cell or bipolar cell. Synapses were formed primarily onto ganglion cells and SACs, with few outputs onto bipolar cells, consistent with findings that bipolar cell terminals are not directionally tuned[26–28]. ON SACs devoted a higher fraction of outputs to ganglion cells than did OFF SACs, possibly because ON SACs provide inputs to both ON–OFF DSGCs and ON DSGCs.

Proximal excitation enhances SAC direction selectivity

Our anatomical data indicate that bipolar cell inputs are restricted to the proximal two-thirds of SAC dendrites and SAC inputs are restricted to the proximal third. Next, we combined computational modelling and physiological imaging to examine how this connectivity pattern affects response properties of SAC dendrites.

We based a single-cell SAC model on an existing passive model[9] and incorporated measured dendritic diameters and active conductances along the dendrites such that the dendrites and soma both preferred centrifugal motion (Extended Data Fig. 7 and Extended Data Table 1)[6]. We then constructed a network model comprising one central SAC and six surrounding SACs (Fig. 4a). SAC–SAC synapses were formed when a presynaptic dendrite came within a defined distance of a postsynaptic cell. The inter-soma distance (145 μm) was set to reproduce the relative angle distributions observed anatomically (Extended Data Fig. 5) and the radial distribution of inhibitory synapses (Fig. 4b, upper panel). We then measured the direction selectivity index (see Methods) at a distal dendritic location (the region of interest (ROI*), Fig. 4d) on the central SAC (Fig. 4a, boxed region).

In response to moving bar stimuli, the ROI* preferred centrifugal motion compared to centripetal motion, as expected (Fig. 4d). During centrifugal motion, depolarization of the dendritic tips preceded inhibition from neighbouring SACs. During centripetal motion, inhibition preceded excitation and limited depolarization of the ROI*. We then modified the model to test whether the spatial separation between excitatory inputs and SAC outputs is important for direction selectivity. When bipolar cell inputs were uniformly distributed along SAC dendrites, thereby overlapping with outputs, the ROI* preferred centripetal over centrifugal motion (Fig. 4e). Bipolar cell inputs on distal tips increased surround inhibition during centrifugal motion and caused excitation to lead inhibition during centripetal motion, thereby reducing direction selectivity. This result suggests that restricting excitation to the proximal two-thirds of SAC dendrites establishes a temporal pattern of excitation and inhibition that enhances preference for centrifugal motion.

Inhibition shapes velocity tuning

When we simulated rabbit-like connectivity by increasing the inter-soma distances (200 μm) to generate distal SAC–SAC contacts (Fig. 4b, lower graph, 4c), the model still exhibited centrifugal preference. The most obvious distinction between the mouse and rabbit eye is a fivefold difference in diameter (Extended Data Fig. 8a)[29,30]. Consequently, a 1° visual angle subtends 30 μm on the mouse retina and 150 μm on the rabbit retina. Mouse and rabbit DSGCs respond to similar angular velocities[31,32] (Extended Data Fig. 8c), suggesting that SACs in both species are also tuned to similar angular velocities. This translates to different linear velocities: 10° s^{-1} motion corresponds to 1,500 μm s^{-1} across rabbit retina, but just 300 μm s^{-1} across mouse retina (Extended Data Fig. 8b).

Both the mouse and rabbit SAC models exhibited direction selectivity at linear velocities above 500 μm s^{-1} (Fig. 4g). At lower velocities,

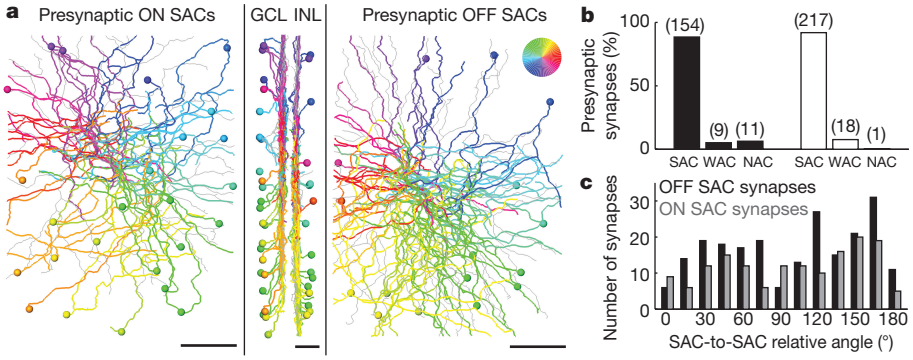


Figure 3 | Inhibitory inputs to mouse SACs. **a**, SAC dendrites presynaptic to an ON and OFF SAC, colour-coded by the orientation of the presynaptic soma relative to the synaptic contact. A total of 33% ($n = 30$) of OFF dendrites and 45% ($n = 30$) of ON dendrites traced back to somas within the data set, corresponding to inter-soma distances between connected SACs of 98.5 ± 35.9 μm (OFF, mean ± s.d.) and 113.4 ± 37.0 μm (ON, mean ± s.d.), consistent with the spacing of connected SACs based on paired recordings in adult mice[43]. **b**, Input synapses originating from different amacrine cells. **c**, Histogram of relative angle (θ) between each presynaptic and postsynaptic SAC dendrite for OFF ($n = 217$, black) and ON ($n = 154$, grey) SAC synapses. Scale bar, 50 μm. NAC, narrow-field amacrine cells; WAC, wide-field amacrine cells.

© 2016 Macmillan Publishers Limited. All rights reserved

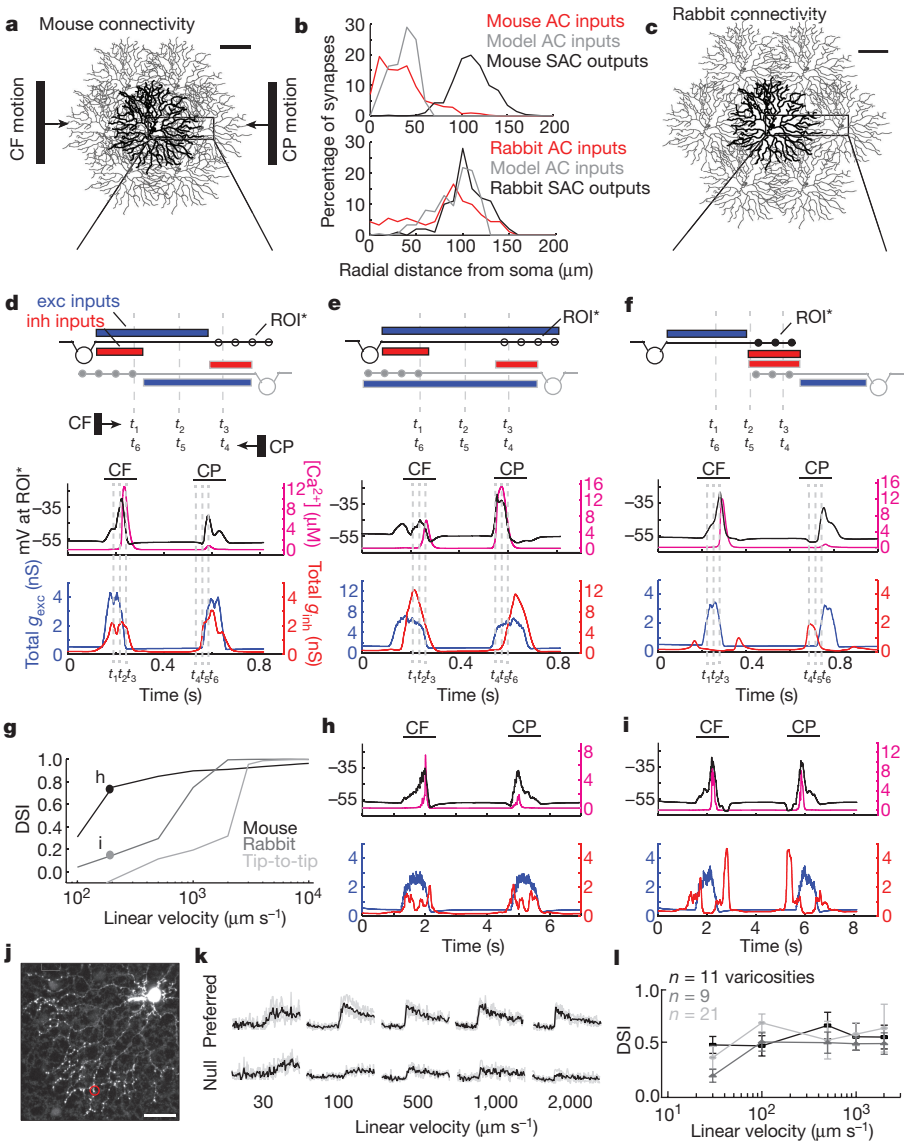


Figure 4 | Functional consequences of SAC network connectivity. **a**, **c**, Compartmental models of mouse (**a**) and rabbit (**c**) networks. **b**, Radial distributions of simulated synapses compared to anatomical reconstructions (rabbit data analysed from ref. 2). **d**, Schematic of mouse connectivity (top) and simulated responses (bottom) to centrifugal (CF) and centripetal (CP) bar stimuli relative to the location ROI*. Bar location at times t_1–t_6 indicated by dashed grey lines. Voltage and calcium responses measured at the ROI*; synaptic conductances measured for the central SAC. **e**, As in **d**, but with bipolar cell inputs distributed uniformly along SAC dendrites. **f**, As in **d**, but incorporating rabbit-like connectivity. **g**, Simulated velocity tuning curves. The direction selectivity index calculated from $[Ca^{2+}]$ at the ROI*. **h**, **i**, Simulated responses at 200 μm s^{-1} for mouse and rabbit models, respectively. **j**, Fluorescence image of an ON SAC filled with OGB1. **k**, Representative Ca^{2+} transients measured at the varicosity highlighted in **j** in response to visual stimuli moving at five different velocities (300% contrast). **l**, Velocity tuning of the direction selectivity index (mean ± s.d.) in $n = 41$ SAC varicosities measured from $n = 3$ ON SACs. Scale bars, 100 μm (**a**, **c**), 25 μm (**j**).

however, direction selectivity in the rabbit model degraded because surround inhibition and central excitation did not overlap sufficiently in time to inhibit centripetal responses as strongly (Fig. 4i). The reduced direction selectivity at lower velocities is consistent with velocity tuning measured in rabbit DSGCs (Extended Data Fig. 8)[31]. By contrast, the mouse model remained direction selective down to 100 μm s^{-1}. The greater spatial overlap of synaptic inputs from neighbouring SACs and bipolar cells in mouse enabled inhibition to coincide with excitation at lower linear velocities during centripetal motion (Fig. 4h). Increasing SAC inter-soma distances to 250 μm, generating tip-to-tip connectivity, further shifted the tuning curve to higher velocities (Fig. 4g).

We tested the prediction from the model by performing two-photon laser scanning microscopy[33] of dendritic calcium from mouse SACs filled with OGB1 in whole-mount retinas[34] (Fig. 4j). Bars of light were swept across SAC receptive fields in eight equally spaced directions at linear velocities ranging from 30–2,000 μm s^{-1}; direction selectivity was calculated from calcium transients measured at individual distal varicosities. As the model predicted, mouse SACs remained directionally selective down to at least 100 μm s^{-1} (Fig. 4k, l). These results suggest that SAC circuitry has adapted to conserve angular velocity tuning across species.

SAC–SAC inhibition expands contrast range

To encode naturalistic stimuli effectively, SACs must also remain directionally selective over a wide contrast range[24,35], a feature predicted by our model (Fig. 5a, b). Simulations suggested that broad contrast tuning requires SAC–SAC inhibition: at high contrasts, blocking inhibition dramatically reduced direction selectivity in simulated SACs due to saturation of postsynaptic responses to both centrifugal and centripetal stimuli (Fig. 5c).

We tested these predictions by imaging SAC dendritic responses to directional motion at different visual contrasts (Fig. 5d–g). Consistent

© 2016 Macmillan Publishers Limited. All rights reserved

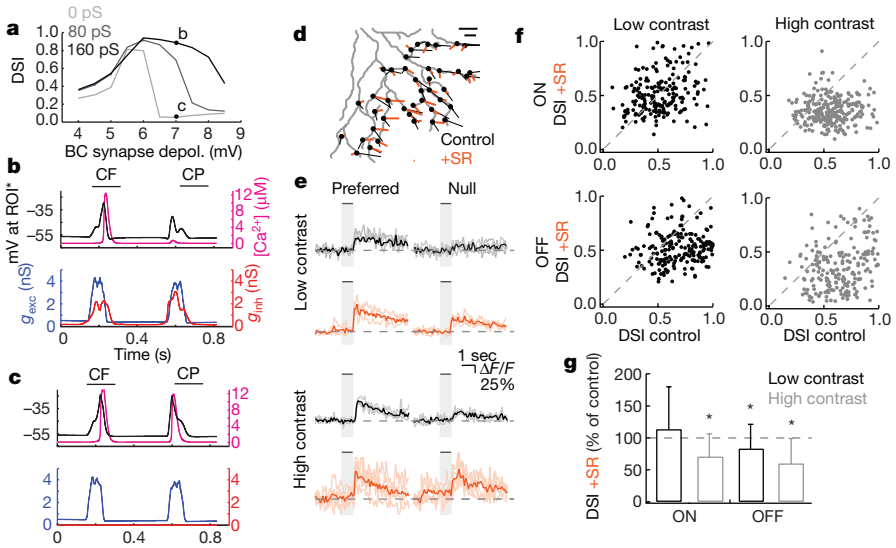


Figure 5 | Contrast dependence of SAC to SAC inhibition. a–c, Contrast tuning curve of mouse network model in response to a bar stimulus. Increasing contrast was simulated with stronger bipolar cell depolarization. The maximal conductance of inhibitory synapses in the model was varied. **d**, Directional tuning of individual SAC varicosities. Vectors indicate preferred direction and the direction selectivity index magnitude. Scale bars, 10 μm (upper), 0.5 direction selectivity index (lower). **e**, Representative Ca^{2+} transients from individual varicosities under low (100%) and high (300%) contrast (SR: SR95531). **f**, The direction selectivity index (DSI) of individual varicosities for ON (left: $n = 201$ ROIs from $n = 6$ cells; right: $n = 261$ ROIs from $n = 10$ cells) and OFF (left: $n = 193$ ROIs from $n = 4$ cells; right: $n = 197$ ROIs from $n = 9$ cells) SACs. **g**, The direction selectivity index (mean ± s.d.) following SR95531 application as a fraction of control (paired t-test, Bonferroni correction, $^*P < 0.001$).

with the model, SACs remained directionally selective over different contrast levels and blocking SAC–SAC inhibition with a $GABA_A$ receptor ($GABA_AR$) antagonist, SR95531 (25 μM), significantly reduced direction selectivity, particularly in response to high-contrast stimuli (Fig. 5g).

Inhibition shapes SAC receptive fields

In rabbit retina, most SAC–SAC connections occur between distal dendrites (Extended Data Fig. 1f)[2]; consequently, direction selectivity for stimuli restricted to a SAC's central receptive field relies primarily upon intrinsic dendritic conductances rather than network inhibition[6,7]. In the mouse retina, we found that SACs receive SAC inputs exclusively on their proximal dendrites (Fig. 1), suggesting that direction selectivity within the central receptive field may rely on inhibition from neighbouring SACs.

We explored this first in our mouse network model using a radially expanding or contracting ('bullseye') stimulus described previously (Fig. 6a, b)[6]. The model exhibited strong centrifugal direction selectivity in response to the bullseye stimulus with inhibition intact, because proximal inhibitory synapses became activated by centrally restricted stimuli (Fig. 6c). Removing inhibition reduced centrifugal direction selectivity over a range of simulated contrasts (Fig. 6c, d). We tested the prediction from the model by imaging dendritic calcium signals evoked by bullseye stimuli restricted to the SAC dendritic arbor (Fig. 6e). Blocking inhibition with SR95531 significantly reduced directional selectivity (Fig. 6f, g), as predicted. SR95531 may also influence presynaptic inhibition of bipolar cell terminals, potentially disrupting bipolar-cell-type-specific release kinetics. If this were the case, however, dendrite autonomous rabbit SAC directional selectivity should also be reduced by SR95531, in contrast to previous reports[6,7].

Discussion

When reconstructing wiring diagrams, an important question is what level of detail is required to understand mechanistically how a neuronal circuit performs specific computations[36,37]. Our results indicate that seemingly subtle differences in connectivity—such as whether cells receive inputs on proximal versus distal dendrites—can substantially influence neural coding and circuit behaviour. We found that segregating excitatory inputs from synaptic outputs along SAC dendrites helps establish strong centrifugal direction selectivity in a network model of SAC connectivity (Fig. 4e, also see ref. 38). More importantly, comparing wiring diagrams across species revealed a previously unrecognized connectivity difference in direction selectivity circuits of the mouse and rabbit retina (Fig. 1 and Extended Data Fig. 1).

The two species exhibit comparable average SAC dendritic diameters and coverage factors[39–42], suggesting that mouse and rabbit SAC networks theoretically could have been wired similarly. We found instead that the locus of presynaptic inhibition on SACs alters the linear velocity tuning of SAC direction selectivity to compensate for eye size difference and conserve angular velocity tuning across the two species

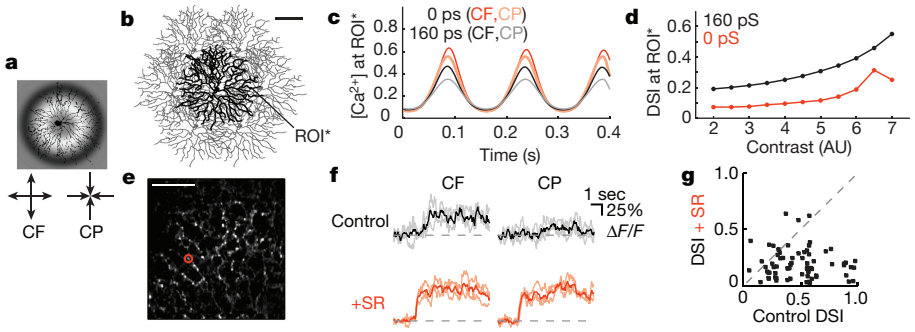


Figure 6 | Receptive field structure of mouse SACs. a, **b**, The mouse network model (**b**) was activated with a bullseye stimulus (**a**) centred on and restricted to the diameter of the central SAC and expanded or contracted to elicit centrifugal or centripetal motion. **c**, Simulated dendritic $[Ca^{2+}]$ at the ROI* in response to centrifugal and centripetal bullseyes (6.7 Hz, 150 μm period, 0.05 AU contrast) with inhibition intact (black, grey) or blocked (orange, peach). **d**, The direction selectivity index versus simulated contrast. **e**, Fluorescence image of OGB1-filled SAC. **f**, Representative dendritic Ca^{2+} transients recorded in response to centrifugal and centripetal bullseye stimuli (2 Hz, 140 μm period, 90% contrast). Responses from the ROI in **e**. **g**, Scatter plot of $n = 74$ ROIs from $n = 5$ ON SACs. SR95531 application significantly decreased the direction selectivity index from 0.46 ± 0.24 (mean ± s.d.) to 0.17 ± 0.14 (paired t-test, $P = 2 \times 10^{-13}$). Scale bars, 100 μm (**b**), 25 μm (**e**).

© 2016 Macmillan Publishers Limited. All rights reserved

(Fig. 4). Inhibition among SACs also extended their contrast tuning range (Fig. 5): removing inhibition reduced SAC directional selectivity at high-stimulus contrasts, potentially rendering postsynaptic DSGCs blind to directional motion. Proximal inhibition also altered the receptive field structure of mouse SACs compared to previous reports of rabbit SACs (Fig. 6)[6,7].

Our simulations effectively guided our physiological experiments, but they underepresented the extensive connectivity of SACs, which actually receive inputs from dozens of neighbouring SACs (Fig. 3). The model also neglects inhibitory inputs to SACs from wide-field amacrine cells and narrow-field amacrine cells and detailed features of the presynaptic bipolar circuitry, important elements to incorporate in future simulations. Other visual stimulus features (for example, size, shape, spatial frequency) also remain to be explored. Nevertheless, the present study exemplifies how connectomic mapping, computational modelling and cellular physiology complement each other to provide new insights into neuronal circuit computations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 December 2015; accepted 27 May 2016.
Published online 22 June 2016.

1. Barlow, H. B., Hill, R. M. & Levick, W. R. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol.* **173,** 377–407 (1964).
2. Famiglietti, E. V. Synaptic organization of starburst amacrine cells in rabbit retina: analysis of serial thin sections by electron microscopy and graphic reconstruction. *J. Comp. Neurol.* **309,** 40–70 (1991).
3. Vaney, D. I., Collin, S. P. & Young, H. M. in *Neurobiology of the Inner Retina* (eds Weiler R. & Osborne N. N.) 157–168 (Springer, 1989).
4. Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471,** 183–188 (2011).
5. Wei, W., Hamby, A. M., Zhou, K. & Feller, M. B. Development of asymmetric inhibition underlying direction selectivity in the retina. *Nature* **469,** 402–406 (2011).
6. Hausselt, S. E., Euler, T., Detwiler, P. B. & Denk, W. A dendrite-autonomous mechanism for direction selectivity in retinal starburst amacrine cells. *PLoS Biol.* **5,** e185 (2007).
7. Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418,** 845–852 (2002).
8. Lee, S. & Zhou, Z. J. The synaptic mechanism of direction selectivity in distal processes of starburst amacrine cells. *Neuron* **51,** 787–799 (2006).
9. Tukker, J. J., Taylor, W. R. & Smith, R. G. Direction selectivity in a model of the starburst amacrine cell. *Vis. Neurosci.* **21,** 611–625 (2004).
10. Gavrikov, K. E., Dmitriev, A. V., Keyser, K. T. & Mangel, S. C. Cation–chloride cotransporters mediate neural computation in the retina. *Proc. Natl Acad. Sci. USA* **100,** 16047–16052 (2003).
11. Oesch, N. W. & Taylor, W. R. Tetrodotoxin-resistant sodium channels contribute to directional responses in starburst amacrine cells. *PLoS One* **5,** e12447 (2010).
12. Kim, J. S. *et al.* Space-time wiring specificity supports direction selectivity in the retina. *Nature* **509,** 331–336 (2014).
13. Taylor, W. R. & Smith, R. G. The role of starburst amacrine cells in visual signal processing. *Vis. Neurosci.* **29,** 73–81 (2012).
14. Münch, T. A. & Werblin, F. S. Symmetric interactions within a homogeneous starburst cell network can lead to robust asymmetries in dendrites of starburst amacrine cells. *J. Neurophysiol.* **96,** 471–477 (2006).
15. Enciso, G. A. *et al.* A model of direction selectivity in the starburst amacrine cell network. *J. Comput. Neurosci.* **28,** 567–578 (2010).
16. Millar, T. J. & Morgan, I. G. Cholinergic amacrine cells in the rabbit retina synapse onto other cholinergic amacrine cells. *Neurosci. Lett.* **74,** 281–285 (1987).
17. Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2,** e329 (2004).
18. Dacheux, R. F., Chimento, M. F. & Amthor, F. R. Synaptic input to the on–off directionally selective ganglion cell in the rabbit retina. *J. Comp. Neurol.* **456,** 267–278 (2003).
19. Roska, B. & Werblin, F. Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature* **410,** 583–587 (2001).
20. Baden, T., Berens, P., Bethge, M. & Euler, T. Spikes in mammalian bipolar cells support temporal layering of the inner retina. *Curr. Biol.* **23,** 48–52 (2013).
21. Borghuis, B. G., Marvin, J. S., Looger, L. L. & Demb, J. B. Two-photon imaging of nonlinear glutamate release dynamics at bipolar cell synapses in the mouse retina. *J. Neurosci.* **33,** 10972–10985 (2013).
22. Greene, M. J., Kim, J. S. & Seung, H. S. Analogous convergence of sustained and transient inputs in parallel on and off pathways for retinal motion computation. *Cell Rep.* **14,** 1892–1900 (2016).
23. Ichinose, T., Fyk-Kolodziej, B. & Cohn, J. Roles of ON cone bipolar cell subtypes in temporal coding in the mouse retina. *J. Neurosci.* **34,** 8761–8771 (2014).
24. Hoggarth, A. *et al.* Specific wiring of distinct amacrine cells in the directionally selective retinal circuit permits independent coding of direction and size. *Neuron* **86,** 276–291 (2015).
25. Ishii, T. & Kaneda, M. ON-pathway-dominant glycinergic regulation of cholinergic amacrine cells in the mouse retina. *J. Physiol.* **592,** 4235–4245 (2014).
26. Park, S. J. H., Kim, I.-J., Looger, L. L., Demb, J. B. & Borghuis, B. G. Excitatory synaptic inputs to mouse On-Off direction-selective retinal ganglion cells lack direction tuning. *J. Neurosci.* **34,** 3976–3981 (2014).
27. Yonehara, K. *et al.* The first stage of cardinal direction selectivity is localized to the dendrites of retinal ganglion cells. *Neuron* **79,** 1078–1085 (2013).
28. Chen, M., Lee, S., Park, S. J., Looger, L. L. & Zhou, Z. J. Receptive field properties of bipolar cell axon terminals in direction-selective sublaminas of the mouse retina. *J. Neurophysiol.* **112,** 1950–1962 (2014).
29. Bozkir, G., Bozkir, M., Dogan, H., Aycan, K. & Güler, B. Measurements of axial length and radius of corneal curvature in the rabbit eye. *Acta Med. Okayama* **51,** 9–11 (1997).
30. Park, H. *et al.* Assessment of axial length measurements in mouse eyes. *Optometry Vision Sci.* **89,** 296–303 (2012).
31. Chan, Y. C. & Chiao, C. C. Effect of visual experience on the maturation of ON–OFF direction selective ganglion cells in the rabbit retina. *Vision Res.* **48,** 2466–2475 (2008).
32. Weng, S., Sun, W. & He, S. Identification of ON–OFF direction-selective ganglion cells in the mouse retina. *J. Physiol. (Lond.)* **562,** 915–923 (2005).
33. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248,** 73–76 (1990).
34. Euler, T. *et al.* Eyecup scope—optical recordings of light stimulus-evoked fluorescence signals in the retina. *Pflugers Arch.* **457,** 1393–1414 (2009).
35. Grzywacz, N. M. & Amthor, F. R. Robust directional computation in On–Off directionally selective ganglion cells of rabbit retina. *Vis. Neurosci.* **24,** 647–661 (2007).
36. Morgan, J. L. & Lichtman, J. W. Why not connectomics? *Nature Methods* **10,** 494–500 (2013).
37. Denk, W., Briggman, K. L. & Helmstaedter, M. Structural neurobiology: missing link to a mechanistic understanding of neural computation. *Nat. Rev. Neurosci.* **13,** 351–358 (2012).
38. Vlasits, A. L. *et al.* A role for synaptic input distribution in a dendritic computation of motion direction in the retina. *Neuron* **89,** 1317–1330 (2016).
39. Vaney, D. I. 'Coronate' amacrine cells in the rabbit retina have the 'starburst' dendritic morphology. *Proc. R. Soc. Lond. B.* **220,** 501–508 (1984).
40. Tauchi, M. & Masland, R. H. The shape and arrangement of the cholinergic neurons in the rabbit retina. *Proc. R. Soc. Lond. B.* **223,** 101–119 (1984).
41. Pérez De Sevilla Müller, L., Shelley, J. & Weiler, R. Displaced amacrine cells of the mouse retina. *J. Comp. Neurol.* **505,** 177–189 (2007).
42. Keeley, P. W., Whitney, I. E., Raven, M. A. & Reese, B. E. Dendritic spread and functional coverage of starburst amacrine cells. *J. Comp. Neurol.* **505,** 539–546 (2007).
43. Kostadinov, D. & Sanes, J. R. Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. *eLife* **4,** (2015).

Acknowledgements We thank W. Denk for supporting the collection of the serial block-face scanning electron microscopy data in his laboratory. This work was supported by NIH grants EY016607 and EY022070 (RGS), by the NINDS Intramural Research Program (NS003145; J.S.D.) and (NS003133; K.L.B.), the Max-Planck Society (K.L.B.), and the Pew Charitable Trusts (K.L.B.).

Author Contributions H.D., R.G.S. and K.L.B. collected and analysed data; H.D., R.G.S., A.P.-P., J.S.D., and K.L.B. designed the study and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.L.B. (kevin.briggman@nih.gov).

Reviewer Information *Nature* thanks G. Knott and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

No statistical methods were used to predetermine sample size. All *n* values refer to biological replicates. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

EM tissue preparation. An adult wild-type (C57BL/6) mouse (postnatal day 30 (P30)) was anaesthetized with isoflurane (Baxter) inhalation and killed by cervical dislocation. The eyes were enucleated and transferred to a dish containing carboxygenated room-temperature saline, in which the retinas were dissected. All procedures were approved by the local animal care committee and were in accordance with the law of animal experimentation issued by the German Federal Government. We used a commercially available saline (Biometra) that was supplemented with 0.5 mM L-glutamine and carboxygenated (95% O_2/5% CO_2). We hemisected the retina and mounted it on filter paper. The retina was fixed in a solution containing 0.1 M cacodylate buffer, 4% sucrose and 2% glutaraldehyde, pH 7.2 (Serva). The tissue was fixed for 2 h at room temperature and then rinsed in 0.1 M cacodylate buffer plus 4% sucrose overnight. A $1 \times 1\,mm^2$ region of the retina, approximately halfway between the optic disk and the peripheral edge of the retina, was then excised. The tissue was then stained in a solution containing 1% osmium tetroxide, 1.5% potassium ferrocyanide, and 0.15 M cacodylate buffer for 2 h at room temperature. The osmium stain was amplified with 1% thiocarbohydrazide (1 h at 50 °C), and 2% osmium tetroxide (1 h at room temperature). The tissue was then stained with 2% aqueous uranyl acetate for 12 h at room temperature and lead aspartate for 12 h at room temperature. The tissue was dehydrated through an ethanol series (70%, 90%, 100%), transferred to propylene oxide, infiltrated with 50%/50% propylene oxide/Epon Hard, and then 100% Epon Hard. The block was cured at 60 °C for 24 h.

Serial block-face scanning electron microscopy acquisition. The retina (k0725) was cut out of the flat-embedding blocks and re-embedded in Epon Hard, on aluminium stubs for serial block-face scanning electron microscopy, with the retinal plane vertical. The samples were then trimmed to a block face of ~200 μm wide and ~400 μm long. The samples were imaged in a scanning electron microscope with a field-emission cathode (QuantaFEG 200, FEI Company). Back-scattered electrons were detected using a custom-designed detector based on a special silicon diode (AXUV, International Radiation Detectors) combined with a custom-built current amplifier. The incident electron beam had an energy of 2.0 keV and a current of ~110 pA. Images were acquired with a pixel dwell time of 2.5 μs and size of 13.2 nm × 13.2 nm which corresponds to a dose of about 10 electrons per nm^2. Imaging was performed at high vacuum, with the sides of the block evaporation-coated with a 100–200 nm thick layer of gold. The electron microscope was equipped with a custom-made microtome designed by W. Denk that was previously used to collect retinal serial block-face scanning electron microscopy data[4,44]. The section thickness was set to 26 nm. 10,112 consecutive block faces were imaged, resulting in aligned data volumes of 4,992 × 16,000 × 10,112 voxels (1 × 5 mosaic of 3,584 × 3,094 images), corresponding to an approximate spatial volume of $50 \times 210 \times 260\,\mu m^3$. The edges of neighbouring mosaic images overlapped by ~1 μm. The cutting quality degraded during the course of the experiment, meaning the images in the first half of the data volume (approximately the first 5,000 slices) are of higher quality than the second half of the volume. Nevertheless, thin neurites could be manually annotated throughout the volume. The imaged region spanned the inner plexiform layer of the retina and included the ganglion cell layer and part of the inner nuclear layer. Cross-correlation-derived shift vectors between neighbouring mosaic images and consecutive slices were used for a global least-squares fit across all shift vectors to align the data sets off-line to subpixel precision by Fourier shift-based interpolation. The data sets were then split into cubes (128 × 128 × 128 voxels) for viewing in KNOSSOS (http://www.knossostool.org).

Skeleton tracing and contact annotation. Skeletons were traced using KNOSSOS and consisted of nodes and connections between them. Nodes were placed approximately every 250 nm. Synapses were manually identified and annotated within Knossos. All analyses of skeletons were performed using MATLAB (MathWorks).

Modelling. We constructed models of an individual SAC and a network of 7 SACs using the simulation language Neuron-C[45]. We digitized a SAC morphology from a confocal stack of a labelled SAC, but included a multiplicative 'diameter factor' set for each dendritic region based on the dendritic diameters measured from the electron microscopy reconstructions (Extended Data Fig. 7). The SAC network was assembled with an algorithm that synaptically interconnected the SAC dendrites based on their location and orientation. Each SAC typically made a total of 120–250 inhibitory synapses onto its neighbours. The central SAC received about twice the number of inhibitory synapses as the surrounding SACs because of the 'edge effect'. Therefore, to achieve a balance between inhibition in the central SAC and its 6 surrounding SACs, we reduced the conductance of the surround→central inhibitory synapses by 50%. BCs were created in a semi-random pattern and were connected to SACs with ribbon synapses if they were within a criterion distance. Synapses were modelled as Ca^{2+}-driven neurotransmitter release that bound to a postsynaptic channel defined by a ligand-activated Markov sequential-state machine[45,46]. The excitatory conductances were typically set to 230 pS and inhibitory conductances were typically 80–160 pS. Membrane ion channels were defined by a voltage-gated Markov state machine and were placed at densities specified for each region of the cell. See Extended Data Table 1 for biophysical parameters.

The contrast of the stimulus presented to the SAC models was achieved by varying the strength of excitatory input from bipolar cells. This was accomplished by voltage-clamping a presynaptic compartment that represented each bipolar cell according to the spatiotemporal pattern of the stimulus. The presynaptic holding potential in the bipolar cells was just above the threshold for synaptic release, typically approximately −45 mV.

The synaptic connectivity of the SAC output synapses was set automatically by an algorithm based on the orientation of presynaptic and possible candidates for the postsynaptic dendrite. When the orientations of both dendrites were within a specific angular range, a synaptic connection was made. This synaptic placement depended on several other criteria, for example, whether the presynaptic point fit within the allowable spacing and radial distribution on the presynaptic dendrite, and also whether the closest point on the postsynaptic dendrite was within a specified distance. The orientations were computed as the absolute angle from the prospective presynaptic point on the distal dendrite to the soma.

Direction selectivity indices were calculated based on the calcium concentration at a location along a central SAC dendrite using the following equation: DSI = (PD–ND)/PD, where PD is the response in the centrifugal direction and ND is the response in the centripetal direction.

Models were run on an array of 3.2 GHz AMD Opteron CPUs interconnected by Gigabit ethernet, with a total of 220 CPU cores. Simulations of the 7-SAC model took 4–48 h, depending on the model complexity and duration of simulated time. The simulations were run on the Mosix parallel distributed task system under the Linux operating system.

Physiological recordings: tissue and calcium-indicator loading. All physiological animal procedures were conducted in accordance with US National Institutes of Health guidelines, as approved by the National Institute of Neurological Disorders and Stroke Animal Care and Use Committee (ASP 1361). Both male and female adult (P30–P60) ChAT-tdTomato mice were used in the experiments (Jackson Laboratory). The mice were anaesthetized with isoflurane (Baxter) inhalation and killed by cervical dislocation. Retinas were isolated and all subsequent procedures were performed at room temperature in Ames media (Sigma) equilibrated with 95% O_2/5% CO_2. Sharp electrodes were pulled on a P-97 Micropipette Puller (Sutter) with a resistance of 100–150 MOhms. Iontophoresis of Oregon Green 488 BAPTA-1 (OGB1, Life Technologies) into single cells was achieved by applying the buzz function in MultiClamp 700B software at 50 ms pulses (Molecular Devices) while the electrode filled with OGB1 (15 mM in water) was on the cell membrane. Pipettes were withdrawn as soon as cell bodies began to fill, and cells were left to recover for 20–30 min before imaging. To block inhibition, the $GABA_A$ receptor antagonist SR95531 (25 μM, Tocris) was added to the extracellular medium.

Physiological recordings: two-photon microscopy. For two-photon imaging, we used a customized microscope (Sutter Movable Objective Microscope), controlled by ScanImage[47], equipped with through-the-objective light stimulation[34] and two detection channels for fluorescence imaging (green, BP 500-540, and red, BP 575-640; Chroma/Thorlabs). The excitation source was a mode-locked Ti/sapphire laser (Chameleon, Coherent) tuned to 920 nm. The microscope was used to simultaneously visualize ChAT-tdTomato-labelled SACs for single-cell targeting (red channel) and to monitor calcium activity reflected by OGB1 fluorescence changes (green channel). During functional imaging, the scan parameters were 256 × 100 pixels at 10 Hz frame rate. Scanning was triggered by the light stimulation. Field of view during acquisition was 80 μm × 80 μm.

Physiological recordings: light stimulation. Light stimulation was generated by custom-written code in Igor software (Wavemetrics) and 4D Workshop 4 IDE (4D Systems) to control an LCD mask in front of a collimated LED (405 nm, Thorlabs) with a bandpass filter (BP 405, Thorlabs). The stimuli were projected onto the retina through the objective lens (XLUMPlanFL 20 × 0.95 NA water-immersion, Olympus). Stimulus contrast varied between 100–300%, with the 300% stimulus intensity at $\sim 25 \times 10^3$ photons $s^{-1}\,\mu m^{-2}$ on a background intensity of $\sim 6 \times 10^3$ photons $s^{-1}\,\mu m^{-2}$. For the bar stimulus, the bar (400 × 400 μm) moved in one of eight evenly spaced directions at a range of velocities between 0.03–1 mm s^{-1}. The bullseye stimulus was configured as previously described[6]. Each stimulus was repeated 3–5 times and responses were averaged.

Calcium-imaging data analysis. Image stacks were analysed using custom Igor (Wavemetrics) functions. Image segmentation was performed by simple thresholding and ROIs are selected as varicosities along dendrites. Response to each stimulus was calculated as the average $\Delta F/F$ during 1 s after stimulus onset; baseline was determined by measuring the fluorescence before the stimulus. The responses were averaged across stimulus presentations. The direction selectivity index was

© 2016 Macmillan Publishers Limited. All rights reserved

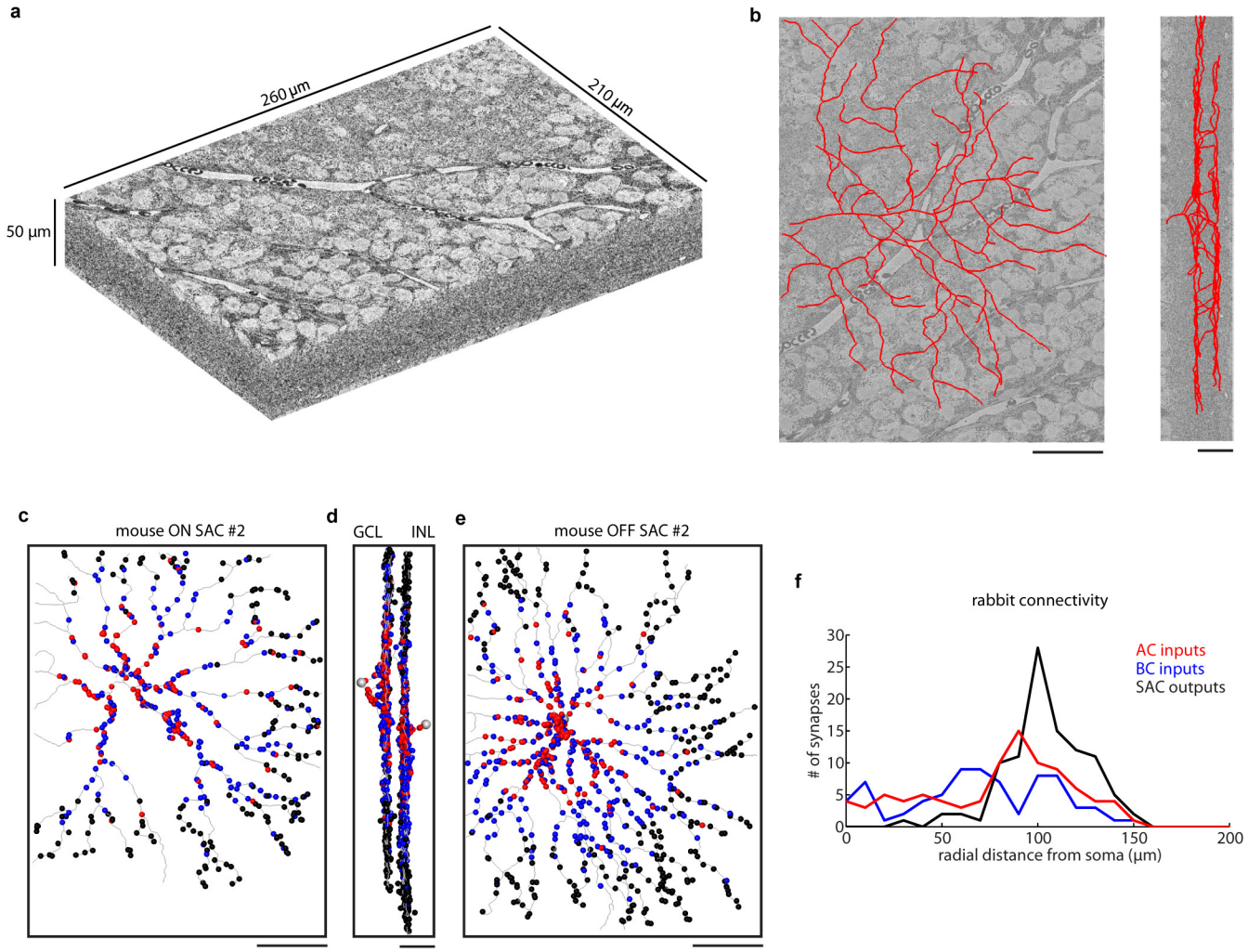
calculated by (PD−ND)/PD, where ND is the null (or centripetal) and PD is the preferred (or centrifugal) response.

Statistical analyses. We included as much of the raw anatomy data as practical in the figures, including neuron and synapse distributions and spatial locations. The identities of neurons presynaptic to SACs were, by definition, blind to the annotator before skeletonization. No reconstructed neurons were excluded from the analysis. For comparing relative angle distributions, we used the non-parametric Kolmogorov–Smirnov test. For dendritic calcium experiments incorporating pharmacology, all measurements were paired (that is, responses at a ROI are reported both before and after drug application). The number of recorded cells was selected to provide typically hundreds of ROIs for comparison and paired *t*-tests were used to assess statistical significance. All samples sizes and statistical test results are reported in the figure legends. Statistical tests were performed in MATLAB or GraphPad.

Code availability. The Neuron-C simulation language that generated the models described above is available at: ftp://retina.anatomy.upenn.edu/pub/rob/nc.tgz. Included in this distribution is the realistic SAC morphology, the 'retsim' retinal circuit simulator that generated the models, and the 'rsbac_stim_plots_vel' script that ran multiple model jobs in parallel.

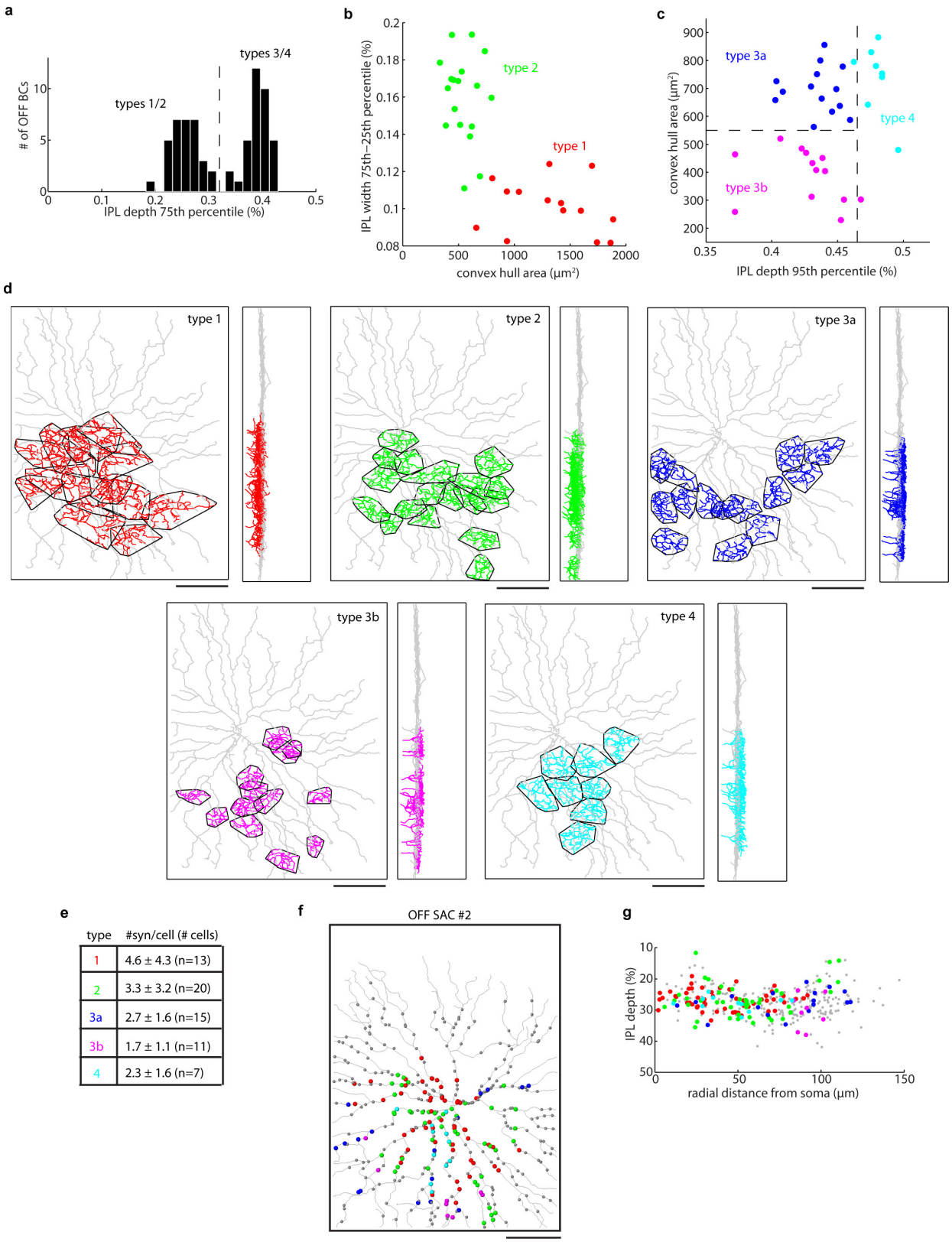
44. Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500,** 168–174 (2013).
45. Smith, R. G. NeuronC: a computational language for investigating functional architecture of neural circuits. *J. Neurosci. Methods* **43,** 83–108 (1992).
46. Schachter, M. J., Oesch, N., Smith, R. G. & Taylor, W. R. Dendritic spikes amplify the synaptic signal to enhance detection of motion in a simulation of the direction-selective ganglion cell. *PLOS Comput. Biol.* **6,** e1000899 (2010).
47. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2,** 13 (2003).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 1 | EM data set, additional SAC reconstructions and rabbit connectivity. a, Conventionally stained serial block-face scanning electron microscopy volume of a mouse retina. **b**, Reconstructed ON-OFF DSGC. **c**–**e**, A second reconstructed ON and OFF SAC with annotated synapses locations. **f**, Annotation of the radial distribution of input and output synapses to and from approximately one-half of an OFF SAC dendritic arbor in rabbit retina. Data analysed from fig. 15 in ref. 2.

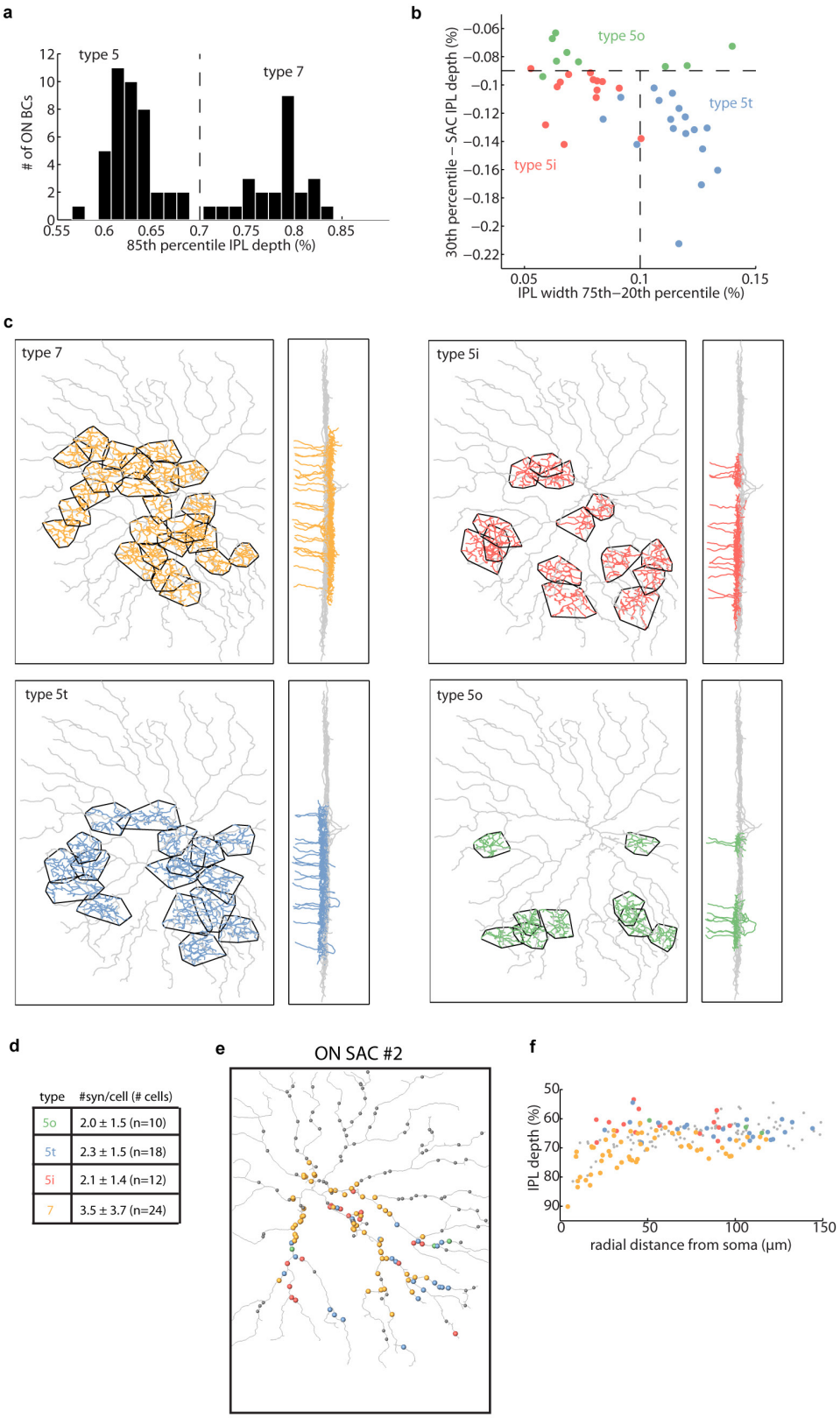
© 2016 Macmillan Publishers Limited. All rights reserved



type	#syn/cell (# cells)
1	4.6 ± 4.3 (n=13)
2	3.3 ± 3.2 (n=20)
3a	2.7 ± 1.6 (n=15)
3b	1.7 ± 1.1 (n=11)
4	2.3 ± 1.6 (n=7)

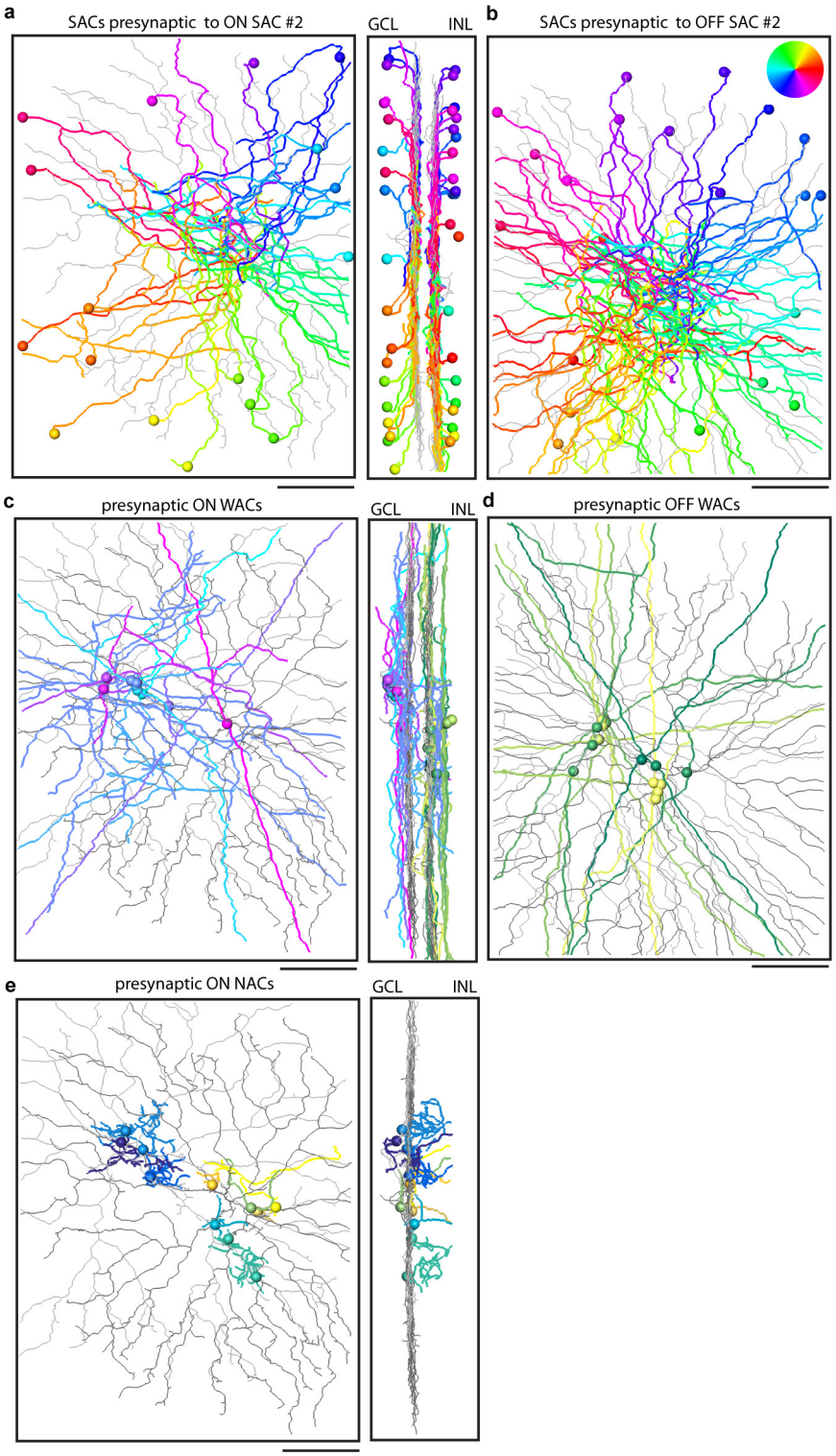
Extended Data Figure 2 | Classification of OFF bipolar cells. a, Types 1/2 and types 3/4 separated by IPL depth. **b**, Types 1 and 2 separate by stratification width and axonal arborization area (convex hull). **c**, Types 3a, 3b and 4 separate by stratification depth and axonal arborization area. **d**, Mosaic patterns and stratification profiles of OFF bipolar cells. **e**, The number of synapses (mean ± s.d.) each bipolar cell, by type, formed with each SAC. **f**, Location of bipolar cell synapses onto a second OFF SAC, colour-coded by bipolar cell type. **g**, The IPL depth of each synapse versus the radial distance relative to the soma.

© 2016 Macmillan Publishers Limited. All rights reserved



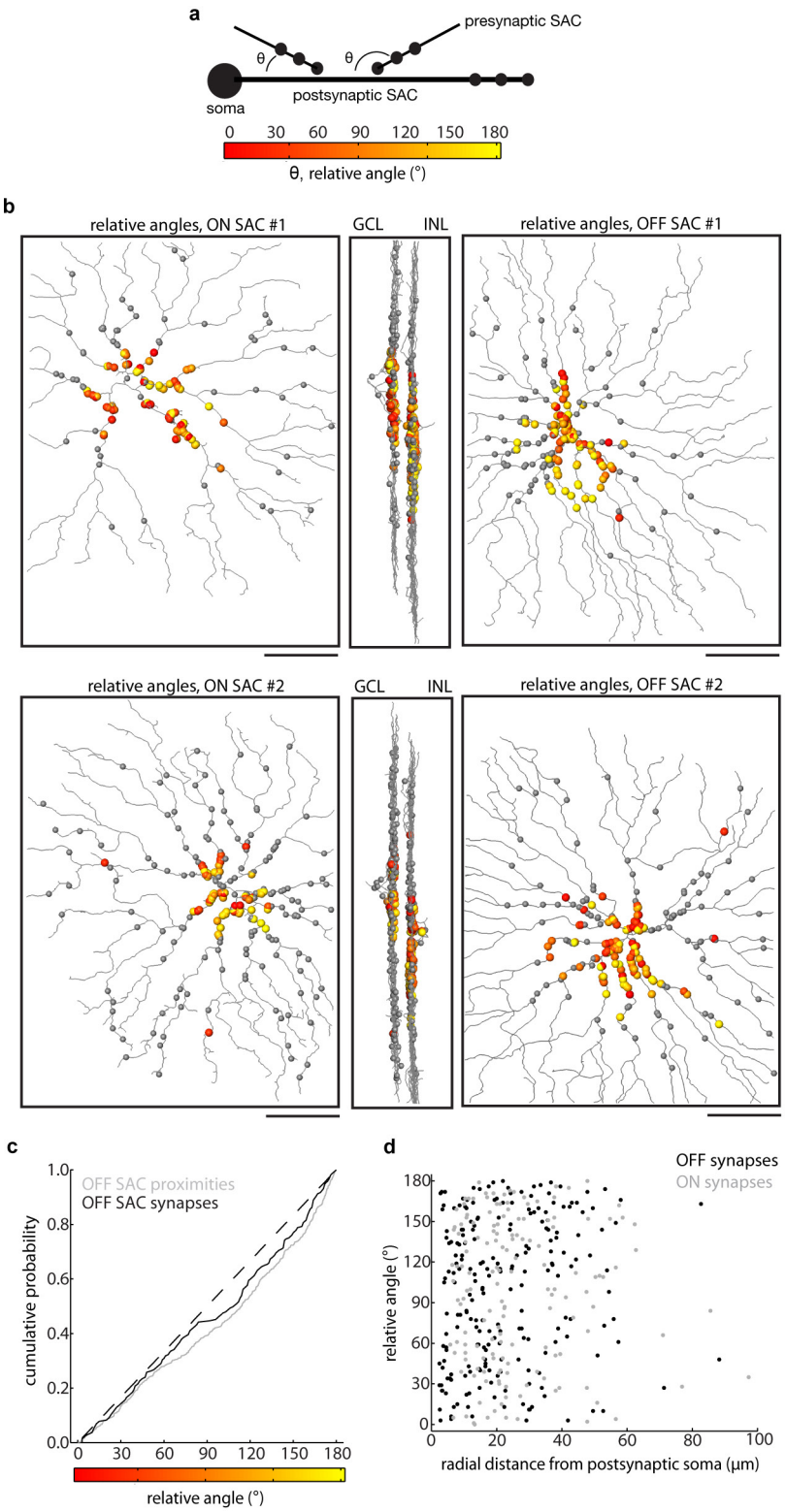
Extended Data Figure 3 | Classification of ON bipolar cells. a, Type 5 and type 7 biploar cells separated by IPL depth. **b**, Types 5o (outer), 5t (thick) and 5i (inner) further subdivide based on IPL depth and stratification width. **c**, Mosaic patterns and stratification profiles of ON bipolar cells. **d**, Summary of the number of synapses (mean ± s.d.) each bipolar cell, by type, formed with each SAC. **e**, Location of bipolar cell synapses onto a second ON SAC, colour-coded by bipolar cell type. **f**, The IPL depth of each synapse versus the radial distance relative to the soma.

© 2016 Macmillan Publishers Limited. All rights reserved



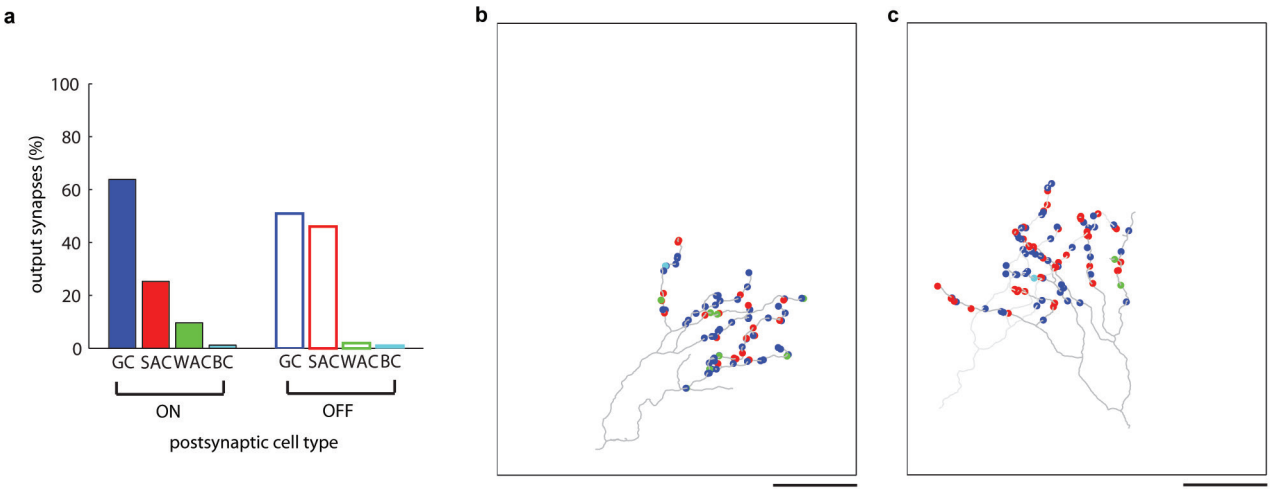
Extended Data Figure 4 | Amacrine cell types presynaptic to SACs. a, **b**, SACs presynaptic to the second pair of mouse SACs colour-coded by absolute orientation. **c**, **d**, Wide-field amacrine cells presynaptic to SACs. **e**, Narrow-field amacrine cells presynaptic to ON SACs.

© 2016 Macmillan Publishers Limited. All rights reserved



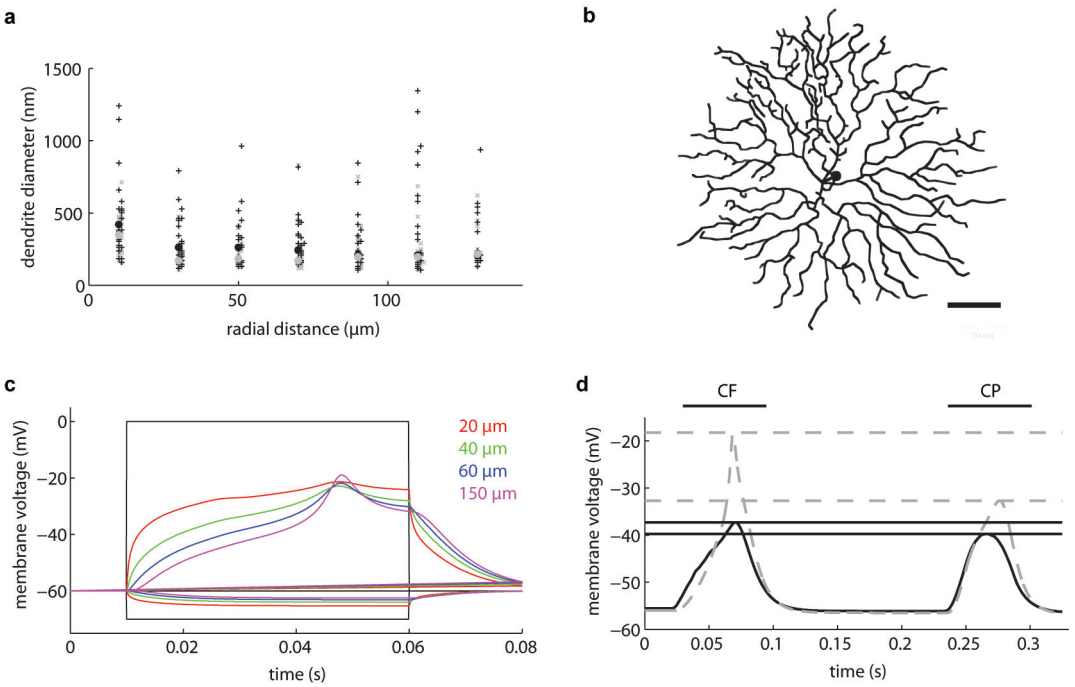
Extended Data Figure 5 | Relative angles between presynaptic and postsynaptic SAC dendrites. a, Schematic of the relative angle measurement: parallel wiring = 0°, anti-parallel wiring = 180°. **b**, Locations of SAC input synapses colour-coded by relative angle. Grey locations indicate AC synapses that were not analysed. **c**, Cumulative distributions of the relative angles between each presynaptic and postsynaptic OFF SAC dendrite for synapses (black) and proximities (grey). Dashed line indicates a uniform distribution. **d**, Relative angle for each synapse was uncorrelated with the radial distance from the postsynaptic somas ($r = 0.07$, $P = 0.16$). Scale bar, 50 μm.

© 2016 Macmillan Publishers Limited. All rights reserved



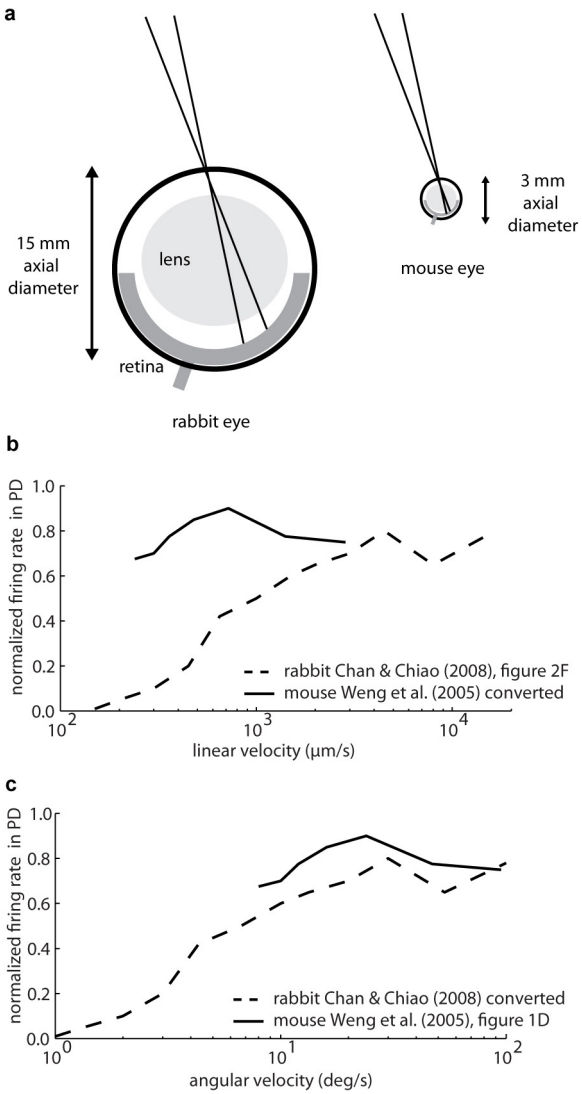
Extended Data Figure 6 | Identities of neurons postsynaptic to SAC output synapses. a, Percentage of output synapses formed with different postsynaptic cell types, colour-coded by postsynaptic cell class: ganglion cells (GC) (blue), SACs (red), bipolar cells (BC) (cyan), and wide-field amacrine cells (WAC) (green). **b**, Locations of 83 annotated output synapses on 1 ON SAC dendrite fragment. **c**, Locations of 110 annotated output synapses on 2 OFF SAC dendrite fragments. Scale bar, 50 μm.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | Single SAC model. a, Dendrite diameters sampled from an ON SAC (grey) and an OFF SAC (black) at different radial distances from their respective somas. **b**, Single SAC morphology used in all simulations. **c**, Somatic voltage clamp simulation showed poor space clamp of even proximal dendrites. Voltage traces measured at a different distances (20–150 μm) from the soma. **d**, Somatic (solid line) and distal dendrite (dashed line) voltage time series in response to an annulus moving centrifugally or centripetally. The addition of active conductances to SAC dendrites (see Extended Data Table 1) rendered somatic voltage recordings directionally selective for centrifugal compared to centripetal stimulation, consistent with electrophysiological measurements. Scale bar, 50 μm.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | Velocity tuning of rabbit and mouse direction selectivity circuits. a, Schematic of the difference in axial diameters and subtended angle on the retina of rabbit and mouse eyes. **b**, Linear velocity tuning curves from rabbit and mouse ON–OFF DSGCs. **c**, Angular velocity tuning curves from rabbit and mouse ON–OFF DSGCs. Data analysed from fig. 2F of ref. 31 and fig. 1D of ref. 32.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Table of biophysical parameters used in model SACs

a

Biophysical parameters for SAC model		
Rm (Ω-cm^2)		10,000
Ri (Ω-cm^2)		75
NaV1.8 channel density (S/cm^2)	soma:	0
	proximal 1/3:	0
	medial 1/3:	$3e^{-3}$
	distal 1/3:	$3e^{-3}$
Kdr channel density (S/cm^2)	soma	$3e^{-3}$
	proximal 1/3:	$2e^{-3}$
	medial 1/3:	$2e^{-3}$
	distal 1/3:	$2e^{-3}$
L-type Ca2+ channel density (S/cm^2)	soma	0
	proximal 1/3:	0
	medial 1/3:	$1e^{-3}$
	distal 1/3:	$1e^{-3}$

© 2016 Macmillan Publishers Limited. All rights reserved

ARTICLE

doi:10.1038/nature18590

Pore-forming activity and structural autoinhibition of the gasdermin family

Jingjin Ding[1,2], Kun Wang[2], Wang Liu[2], Yang She[1,2], Qi Sun[2], Jianjin Shi[2], Hanzi Sun[2], Da-Cheng Wang[1,3] & Feng Shao[1,2,4]

Inflammatory caspases cleave the gasdermin D (GSDMD) protein to trigger pyroptosis, a lytic form of cell death that is crucial for immune defences and diseases. GSDMD contains a functionally important gasdermin-N domain that is shared in the gasdermin family. The functional mechanism of action of gasdermin proteins is unknown. Here we show that the gasdermin-N domains of the gasdermin proteins GSDMD, GSDMA3 and GSDMA can bind membrane lipids, phosphoinositides and cardiolipin, and exhibit membrane-disrupting cytotoxicity in mammalian cells and artificially transformed bacteria. Gasdermin-N moved to the plasma membrane during pyroptosis. Purified gasdermin-N efficiently lysed phosphoinositide/cardiolipin-containing liposomes and formed pores on membranes made of artificial or natural phospholipid mixtures. Most gasdermin pores had an inner diameter of 10–14 nm and contained 16 symmetric protomers. The crystal structure of GSDMA3 showed an autoinhibited two-domain architecture that is conserved in the gasdermin family. Structure-guided mutagenesis demonstrated that the liposome-leakage and pore-forming activities of the gasdermin-N domain are required for pyroptosis. These findings reveal the mechanism for pyroptosis and provide insights into the roles of the gasdermin family in necrosis, immunity and diseases.

Pyroptosis is critical for host defences against infection and danger signals, and excessive pyroptosis causes immunological diseases and septic shock. Pyroptosis involves cell swelling and lysis, which causes massive release of cellular contents and thereby triggers strong inflammation[1]. The term pyroptosis was originally used to describe caspase-1-mediated macrophage death[2]. Caspase-1 belongs to the inflammatory caspase group, which also includes murine caspase-11 and its human counterparts caspase-4 and -5 (ref. 3). Unlike caspase-11 (ref. 4), caspase-4 and -5 can also activate pyroptosis in non-monocytotic cells[3]. Caspase-1 acts downstream of the inflammasome complex, which is scaffolded by an Nod-like receptor (NLR) protein, absent in melanoma 2 (AIM2) or pyrin, and recognizes bacteria, other microbes and endogenous threats[5,6]. Caspase-4, -5 and -11 sense[7,8] and are activated by direct binding to lipopolysaccharide (LPS)[3,9]; hyperactivation of these caspases causes septic shock.

Recent studies have identified the GSDMD protein, which critically determines pyroptosis[10,11]. GSDMD is cleaved by all inflammatory caspases between its N-terminal gasdermin-N and C-terminal gasdermin-C domains. This cleavage releases the autoinhibition by gasdermin-C of gasdermin-N, which has intrinsic pyroptosis-inducing activity. The absence of GSDMD does not affect caspase-1 processing of interleukin (IL)-1β, but blocks mature IL-1β secretion, suggesting that pyroptosis is required for noncanonical cytokine secretion[10,11]. Besides GSDMD, the gasdermin family also includes GSDMA, GSDMB, GSDMC, DFNA5 and DFNB59 (refs 12,13). Mice lack GSDMB but have three GSDMA (GSDMA1–3) and four GSDMC (GSDMC1–4) proteins. Other gasdermins are insensitive to inflammatory caspases[10]. Dominant mutations in *Gsdma3* (refs 14–16) and *DFNA5* (or autosomal recessive mutation in *DFNB59*)[17,18] cause alopecia and hyperkeratosis in mice and nonsyndromic hearing loss in humans, respectively. Disease-associated mutants of GSDMA3 and its gasdermin-N domain alone can activate pyroptosis owing to loss of autoinhibiton[10]. Despite the importance of gasdermins in pyroptosis and inflammation, the mechanisms of action of GSDMD and the gasdermin family are unknown.

Cytotoxicity of gasdermin-N from multiple gasdermins

All gasdermins, except for DFNB59, adopt a two-domain architecture. As found in GSDMD and GSDMA3 (ref. 10), the gasdermin-N domains of GSDMA, GSDMB, GSDMC or DFNA5, but not the full-length proteins, induced extensive pyroptosis in human 293T cells (Extended Data Fig. 1a, b). This suggests that gasdermins in general are pyroptosis factors. Expression of the GSDMD gasdermin-N domain (GSDMD-N) is highly toxic to *Escherichia coli* (Extended Data Fig. 1c, d), whereas little cytotoxicity was observed with full-length GSDMD and its GSDMD-C domain (both proteins could be highly expressed in *E. coli*). Similar phenomena were observed with GSDMA, GSDMA3, GSDMC and DFNA5 (Extended Data Fig. 1d). Thus, the gasdermin-N domain has intrinsic cytotoxicity in mammalian cells and its overexpression can also kill bacteria.

Gasdermin-N domains can bind membrane lipids

We hypothesized that the gasdermin-N domains might disrupt membranes to cause pyroptosis. To test this hypothesis, we assayed the binding of recombinant gasdermin-N to membrane lipids. To circumvent the toxicity of gasdermin-N in *E. coli*, a PreScission protease (PPase) site (LEVLFQGP) was engineered into the inter-domain linker in GSDMD, GSDMA and GSDMA3 (it has been shown that PPase cleavage of the engineered GSDMD or GSDMA3 can trigger pyroptosis[10]) and purified full-length gasdermins. Notably, the gasdermin-N and -C domains remained bound together following *in vitro* PPase cleavage. When the noncovalent complexes (N+C) were incubated with liposomes containing phosphatidylcholine as the skeleton lipid, all three gasdermin-N (30–35 kDa) but not gasdermin-C domains (20–25 kDa) were precipitated by liposomes containing 10% or 20% phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P_2) (a major phosphoinositide in the plasma membrane), but not unphosphorylated phosphatidylinositol (Fig. 1a and Extended Data Fig. 2a, b). Neither full-length GSDMD, GSDMA and GSDMA3 nor their gasdermin-C domains bound phosphoinositide. Gasdermin-N could also bind to liposomes containing monophosphorylated (PtdIns3P, PtdIns4P and PtdIns5P),

[1]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, China. [2]National Institute of Biological Sciences, Beijing, 102206, China. [3]Foshan University, Guangdong, 528000, China. [4]National Institute of Biological Sciences, Collaborative Innovation Center for Cancer Medicine, Beijing, 102206, China.

© 2016 Macmillan Publishers Limited. All rights reserved

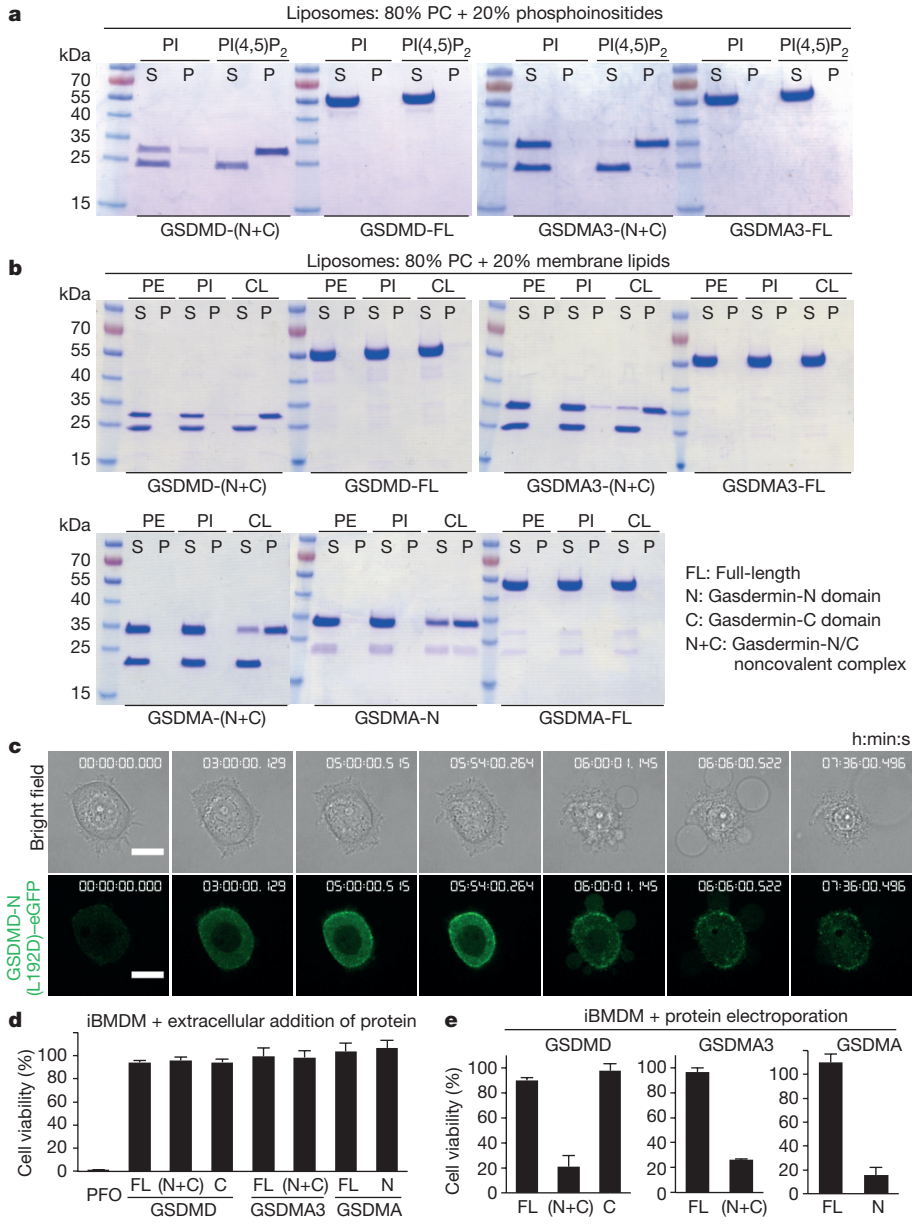


Figure 1 | Lipid binding, biomembrane association and disruption by the gasdermin-N domain. a, **b**, Binding of the gasdermin-N domain to membrane lipids. Purified gasdermin proteins were incubated with liposomes with indicated lipid compositions. After ultracentrifugation, the liposome-free supernatant (S) and the liposome pellet (P) were analysed by SDS–PAGE and Coomassie blue staining. CL, cardiolipin; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PI, phosphatidylinositol. The gasdermin-N and -C noncovalent complex was obtained from inter-domain cleavage of the full-length (FL) protein. **c**, Microscopy imaging of GSDMD-N localization in cells during pyroptosis. GSDMD-N(L192D)–eGFP was stably expressed in HeLa cells under a tetracycline-inducible promoter. Shown are representative time-lapse cell images (brightfield and fluorescence) taken from 4–5 h after doxycycline addition. Scale bar, 15 μm. For videos of two representative cells, see Supplementary Videos 1 and 2. **d**, **e**, Effects of extracellular or intracellular delivery of purified gasdermin proteins on iBMDM cell viability. Equal amounts of indicated gasdermin proteins or PFO were added directly into cell culture medium (**d**) or electroporated into the cytosol (**e**). ATP-based cell viability is expressed as mean ± s.d. from three technical replicates. All data shown are representative of three independent experiments.

bisphosphorylated (PtdIns(3,4)P_2 and PtdIns(3,5)P_2) or triphosphophorylated (PtdIns(3,4,5)P_3) phosphatidylinositols (Extended Data Fig. 2b). Similar PtdIns(4,5)P_2 binding was observed with liposomes made of complicated phospholipid mixtures (phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol and PtdIns(4,5)P_2) that mimic plasma membrane lipid composition (Extended Data Fig. 2c).

Cardiolipin and phosphatidylethanolamine are major bacterial membrane lipids. The gasdermin-N domains of GSDMD, GSDMA and GSDMA3, but not the full-length proteins or their gasdermin-C domains, were efficiently and specifically precipitated by cardiolipin liposomes (Fig. 1b). Reducing cardiolipin concentration from 20% to 10% decreased the binding efficiency (Fig. 1b and Extended Data Fig. 2d). Specific binding of gasdermin-N to cardiolipin, as well as the phosphoinositides, was also evident in the lipid-strip binding assay (Extended Data Fig. 2e). Thus, cardiolipin and phosphoinositide are two membrane lipid targets of gasdermin-N.

For the three gasdermins GSDMD, GSDMA and GSDMA3, only the gasdermin-N domain of GSDMA could be separated from the noncovalent complex by high-salt buffer. The apo-GSDMA-N domain showed similar binding to cardiolipin or phosphoinositide liposomes as the noncovalent complex (Fig. 1b and Extended Data Fig. 2a, c, d), suggesting that lipid binding by GSDMA does not involve gasdermin-C. Moreover, the three gasdermin-N domains bound strongly to cardiolipin liposomes; GSDMA-N showed weaker binding to phosphoinositide liposomes than GSDMD-N or GSDMA3-N despite their comparable pyroptosis-inducing activity (Extended Data Fig. 1a, b). A possible cause of this phenomenon is that artificial liposomes may not exactly recapitulate the complex lipid constituents of biomembranes.

Membrane targeting of gasdermin-N during pyroptosis

We next examined the localization of gasdermin-N during pyroptosis. Extracts of *Gsdmd*$^{-/-}$ immortalized bone marrow macrophages (iBMDMs) stably expressing Flag–GSDMD[10] were sequentially centrifuged (700*g*, 20,000*g* and 100,000*g*). Full-length GSDMD was found exclusively in the supernatant after 100,000*g* centrifugation (S100) (Extended Data Fig. 3a), suggesting a cytosolic localization. When pyroptosis was induced by LPS electroporation, GSDMD was cleaved into GSDMD-N and GSDMD-C; while GSDMD-C remained in S100, a large portion of GSDMD-N was distributed in the P7 heavy-membrane (pellet from 700*g* centrifugation) and P20 light-membrane fractions, which resembles the distribution of LAMP1 (endosome/lysosome),

© 2016 Macmillan Publishers Limited. All rights reserved

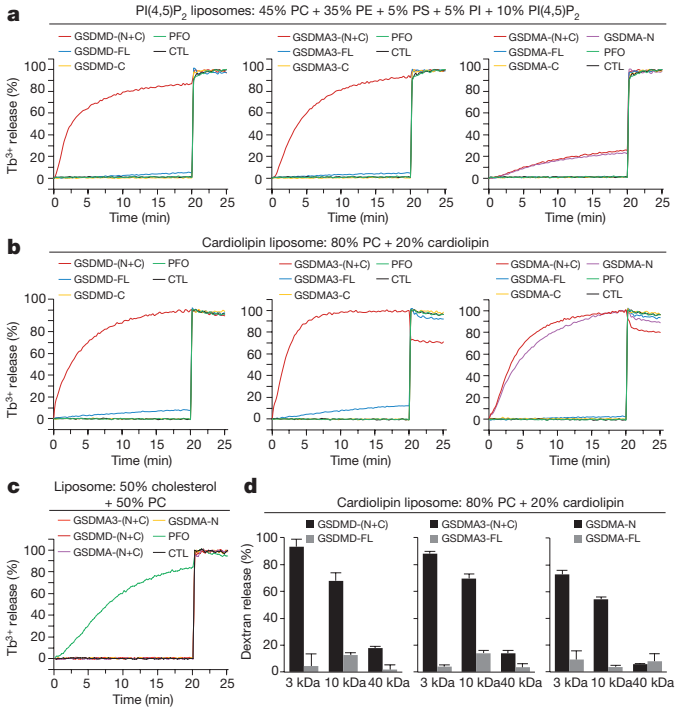


Figure 2 | Liposome-leakage-inducing activity of the gasdermin-N domain. Liposomes with indicated lipid compositions were treated with purified gasdermin proteins or PFO as indicated. **a–c**, Liposome leakage was monitored by measuring 2,6-pyridinedicarboxylic acid (DPA) chelating-induced fluorescence of released Tb^{3+}. Time course of relative Tb^{3+} release is shown. CTL, control. **d**, Different-size fluorescent dextrans were encapsulated into the liposome and the released dextran fluorescence was determined. Triton X-100 treatment was used to achieve 100% liposome leakage. Final leakage of the dextrans is expressed as mean ± s.d. from three technical replicates. All data shown are representative of three independent experiments.

Na, K-ATPase α1 (plasma membrane) and syntaxin 6 (*trans*-Golgi) (Extended Data Fig. 3a). GSDMD-N was also found in the lightest P100 fraction. As expected, actin and mitochondrial proteins (COX4 and TOM20) were found in the S100 and P7 fractions, respectively. The gasdermin-N domain of PPase-cleavable GSDMA showed the same distribution in pyroptotic 293T cells (Extended Data Fig. 3b). Thus, certain gasdermin-N domains moved to heterogeneous membranes during pyroptosis, echoing the binding to various phosphoinositides.

To visualize the membrane targeting, GSDMD-N and GSDMA3-N fused to enhanced green fluorescent protein (eGFP) were inducibly expressed in HeLa cells. The pyroptosis triggered by gasdermin-N expression was too severe and rapid to allow accumulation of fluorescence signals sufficient for detection. This was overcome by a mutation of leucine192 to aspartate (L192D), which slowed down the pyroptosis (see below). GSDMD-N(L192D) showed initial cytoplasmic distribution, and some of it was translocated and accumulated on the plasma membrane as pyroptosis progressed; the cells then developed characteristic swelling bubbles and became ruptured (Fig. 1c and Supplementary Videos 1, 2). Similar results were obtained with GSDMA3-N containing an equivalent L184D mutation (Extended Data Fig. 3c and Supplementary Videos 3, 4).

Biomembrane disruption correlates with lipid binding

Perfringolysin O (PFO) is a cholesterol-targeting pore-forming toxin from *Clostridium perfringens*. Extracellular addition of purified PFO to iBMDMs caused severe cytotoxicity (Fig. 1d). By contrast, the noncovalent complex of cleaved GSDMD or GSDMA3 or the GSDMA-N domain induced cell lysis only when delivered cytosolically but not when administered extracellularly (Fig. 1d, e). Identical results were obtained in 293T cells (Extended Data Fig. 3d, e). These findings are consistent with the presence of cholesterol in the exoplasmic leaflet of the plasma membrane and the localization of phosphoinositides, the targets of gasdermin-N, only in the cytoplasmic leaflet. To perform a similar assay in bacteria, we used protoplasts of Gram-positive *Bacillus megaterium* containing a single cardiolipin-containing membrane. The protoplasts were completely lysed by the GSDMD-(N+C) noncovalent complex, but not by full-length GSDMD or GSDMD-C (Extended Data Fig. 3f, g). PFO also caused no protoplast lysis, consistent with the absence of cholesterol in bacterial membranes. Similar results were obtained with the gasdermin-N domains of GSDMA and GSDMA3 (Extended Data Fig. 3g). Thus, membrane disruption by gasdermin-N correlates well with its lipid-binding properties.

Liposome leakage triggered by the gasdermin-N domain

We then tested the induction of liposome leakage by the gasdermin-N domain. The noncovalent complex of cleaved GSDMD or GSDMA3 caused around 50% leakage of PtdIns(4,5)P_2 liposomes (Extended Data Fig. 4a). The leakage reached 100% when PtdIns(4,5)P_2 was reconstituted into liposomes containing complicated phospholipid mixtures (Fig. 2a). Consistent with the binding data, GSDMA-N induced less liposome leakage (Fig. 2a and Extended Data Fig. 4a); liposome leakage reached about 50% with higher concentrations of GSDMA-N (Extended Data Fig. 4b). All three gasdermins caused nearly 100% leakage of the cardiolipin liposome (Fig. 2b). Full-length gasdermins and gasdermin-C did not lyse either type of liposome (Fig. 2a, b and Extended Data Fig. 4a). The gasdermins had no effect on cholesterol- or phosphatidylethanolamine-reconstituted liposomes (Fig. 2c and Extended Data Fig. 4c). As expected, PFO lysed cholesterol-containing liposomes but not those containing PtdIns(4,5)P_2 or cardiolipin (Fig. 2a–c and Extended Data Fig. 4a). These results are consistent with the finding that PFO but not gasdermin-N could lyse mammalian cells from the outside (Fig. 1d and Extended Data Fig. 3d). When liposomes encapsulating different-size fluorescent dextrans were assayed, the active forms of GSDMD, GSDMA3 and GSDMA could release dextrans with molecular masses of 3 or 10 kDa but not 40 kDa (Fig. 2d). This indicates that items with diameters of 10 nm or less can pass through the presumed pores formed by gasdermin-N.

Oligomerized gasdermin-N forms membrane pores

Full-length GSDMD, GSDMA3 and GSDMA were monomeric in solution (Extended Data Fig. 5a). Upon crosslinking of the GSDMD- or GSDMA3-(N+C) noncovalent complex or the GSDMA-N domain, GSDMD-N remained monomeric and GSDMA3/GSDMA-N showed a low degree of artificial oligomerization (Extended Data Fig. 5b). When crosslinking was performed after liposome incubation, all three gasdermin-N domains appeared as high-order oligomers (Extended Data Fig. 5b). As a control, PFO was converted from monomers into SDS-resistant oligomers after liposome incubation[19]. In LPS-stimulated pyroptotic iBMDMs, membrane-associated GSDMD-N, resulting from caspase-11 cleavage, also formed high-order oligomers, whereas full-length GSDMD and cytosolic GSDMD-N remained monomeric (Extended Data Fig. 5c). Similar results were obtained with the PPase-cleavable GSDMA in pyroptotic 293T cells (Extended Data Fig. 5c).

Negative-stain electron microscopy revealed multiple pores on nearly all cardiolipin liposomes that had been incubated with the GSDMD- or GSDMA3-(N+C) noncovalent complex, but not with full-length GSDMD or GSDMA3 (Fig. 3a). Similar (but fewer) membrane pores were formed on PtdIns(4,5)P_2-containing liposomes (Extended Data Fig. 5d); both intact and severely fragmented liposomes caused by merging of adjacent pores were observed. The pore-forming efficiency was markedly improved by reconstituting PtdIns(4,5)P_2 into liposomes containing the complicated phospholipid mixtures (Fig. 3a), consistent with the finding that these liposomes showed more severe leakage (Fig. 2a and Extended Data Fig. 4a). Furthermore,

© 2016 Macmillan Publishers Limited. All rights reserved

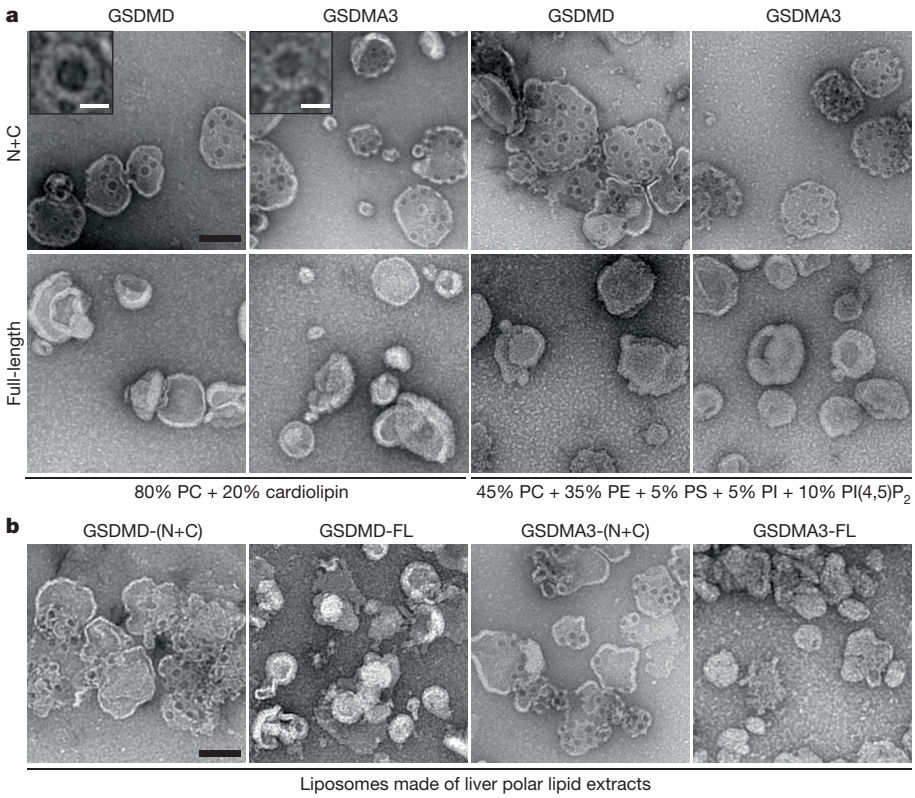


Figure 3 | Membrane pore-forming activity of the gasdermin-N domain. a, **b**, Liposomes with indicated lipid compositions (**a**) or prepared using bovine liver-derived polar lipid extracts (**b**) were treated with indicated gasdermin proteins. Shown are representative negative-stain electron microscopy micrographs of the liposomes (scale bar, 100 nm). Insets in **a**, expanded view of a representative pore (scale bar, 15 nm). All data shown are representative of three independent experiments.

the gasdermin-N domains of GSDMD, GSDMA3 and GSDMA could bind robustly to liposomes made of bovine liver or brain-derived lipid extracts (Extended Data Fig. 2f) and accordingly generated similar pores on the liposomes (Fig. 3b).

Of the GSDMD-induced pores, 60% had inner diameters of 10–16 nm whereas all GSDMA3-induced pores had inner diameters of 10–14 nm (Extended Data Fig. 6a, b). Decreasing the gasdermin concentration by 10-fold affected the number but not the size distribution of the pores (Extended Data Fig. 6a, b). The pore size is consistent with the assessment from the dextran leakage data (Fig. 2d). The wider pore-size range produced by GSDMD, compared with GSDMA3, probably resulted from the less optimal properties of recombinant GSDMD. Previously, the inner diameter of membrane pores in caspase-1-mediated pyroptosis was estimated to be 1.1–2.4 nm, on the basis of the observation that PEG2000 but not PEG200, at an equal molar concentration, can block pyroptosis[20]. We confirmed this finding but further found that increasing the mass concentration of PEG200 to the same as PEG2000 (to ensure equal osmotic potential) could generate the same protective effect (Extended Data Fig. 6c). Considering that gasdermin-N formed similar pores on liposomes made of natural lipid extracts, pores with inner diameters of 10–14 nm are likely to predominate *in vivo*. This pore size can allow the passage of mature IL-1β (also IL-18) and caspase-1, which have diameters of 4.5 and 7.5 nm, respectively. GSDMD-N and GSDMA3-N also formed pores on lipid monolayers; GSDMA3-N pores had a uniform inner diameter of about 14 nm (Extended Data Fig. 6d). Following 2D classification, one class of GSDMA3 pores showed the best protein contrast, and subsequent rotational auto-correlation analysis revealed 16-fold symmetry (Extended Data Fig. 6e). Given the single-layer property of all known pore-forming complexes, these data suggest that the gasdermin-N domain forms a 16-mer pore complex.

Crystal structure of GSDMA3

We determined the 1.90 Å crystal structure of GSDMA3 (Extended Data Table 1 and Extended Data Fig. 7a).The structure is separated into two domains—the presumed gasdermin-N and gasdermin-C domains (Fig. 4a). Gasdermin-N mainly contains an extended twisted β-sheet formed by nine tandem strands (β3–β11) (Fig. 4a and Extended Data Fig. 7b). The N-terminal α1 helix and following β1–β2 hairpin lie in the concave of the β-sheet. Helices α2 and α3 flank the β-sheet at one end. Helix α4 protrudes away from the other end of the β-sheet through two loops to contact the helical gasdermin-C domain. The C domain adopts a compact globular fold covered by a short three-stranded β-sheet (β12–β14). The long loop (residues 234–263) linking the GSDMA3-N and -C domains stretches away from the main body of the structure. A structural homology search[21] revealed no meaningful information about gasdermin-C; the structure of gasdermin-N also showed no convincing similarity to any known proteins, suggesting that it represents a new type of pore-forming protein.

Functional analyses of GSDMA3 autoinhibition

In the structure, the α1 helix and β1–β2 hairpin in GSDMA3-N provide the primary surface for binding to the GSDMA3-C domain (Fig. 4b). F48 and W49 from the hairpin loop are inserted deeply into a groove in GSDMA3-C and encircled by L270, Y344, A348 and A443, forming a hydrophobic core (Fig. 4c). In addition, R43, K44 and T46 from the hairpin have four hydrogen bonds with E273, E277 and D340; α1 also supplies D6 and R13 for hydrogen-bonding with H436 and D433. At the second inter-domain interface, α4 presents its hydrophobic face (L181, L184 and L186) to a non-polar surface formed by α9 and α11 in GSDMA3-C (Fig. 4c). Nine *Gsdma3* mutations that can cause alopecia and hyperkeratosis in mice have been identified[14–16]. Among them, 259RDW (insertion after residue 259 with three mistranslated residues RDW) and 366stop (a premature stop at residue 366) encode truncated GSDMA3 devoid of inter-domain contacts; Y344C, Y344H and A348T are mutations in residues that directly contact GSDMA3-N; T278P and L343P are near the direct-contacting residues; and 412EA (duplication of E411A412) disrupts the α4-binding surface. Consistently, T278P, L343P, Y344C, A348T and 412EA all weaken inter-domain interactions, resulting in constitutive activation of GSDMA3 (ref. 10). Similarly, GSDMA3 L270D, Y344D and A348D exhibited spontaneous pyroptosis-inducing activity (Fig. 4d).

© 2016 Macmillan Publishers Limited. All rights reserved

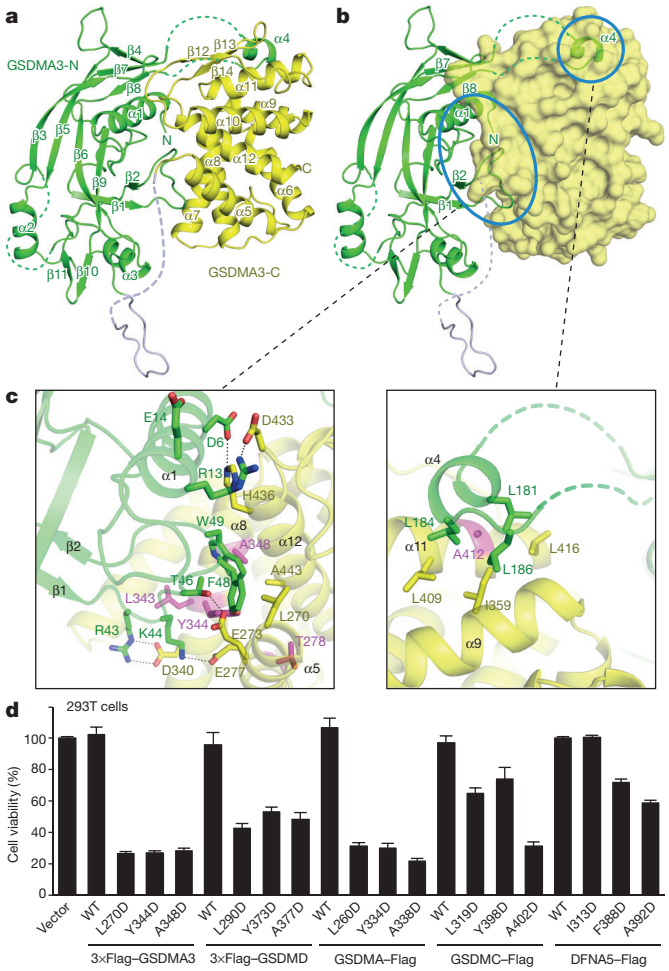


Figure 4 | Crystal structure of GSDMA3 and structural autoinhibition of the gasdermin family. a, **b**, Overall structure of GSDMA3 and the inter-domain interfaces. The gasdermin-N (GSDMA3-N) and gasdermin-C (GSDMA3-C) domains are coloured green and yellow, respectively. Structures of GSDMA3-N (**a**, **b**) and GSDMA3-C (**a**) are shown as cartoon models, and that of GSDMA3-C (**b**) is in a transparent surface scheme. Secondary structure elements are labelled in **a**. The primary and second inter-domain interfaces are highlighted by large and small blue ellipses, respectively (**b**). Disordered loops are indicated by dashed lines. **c**, Close-up view of the autoinhibitory interactions. Left and right, primary and secondary inter-domain interfaces, respectively. Residues involved in the autoinhibitory interactions are labelled and shown as sticks. Point mutations in GSDMA3 that cause alopecia in mice are coloured pink. Dotted lines, hydrogen bonds. **d**, Mutation analyses of the autoinhibitory contacts. Full-length GSDMA3, GSDMD, GSDMA, GSDMC or DFNA5 (wild type (WT) or containing indicated point mutations in their gasdermin-C domains) was transfected into 293T cells. ATP-based cell viability is expressed as mean ± s.d. from three technical replicates; data are representative of three independent experiments.

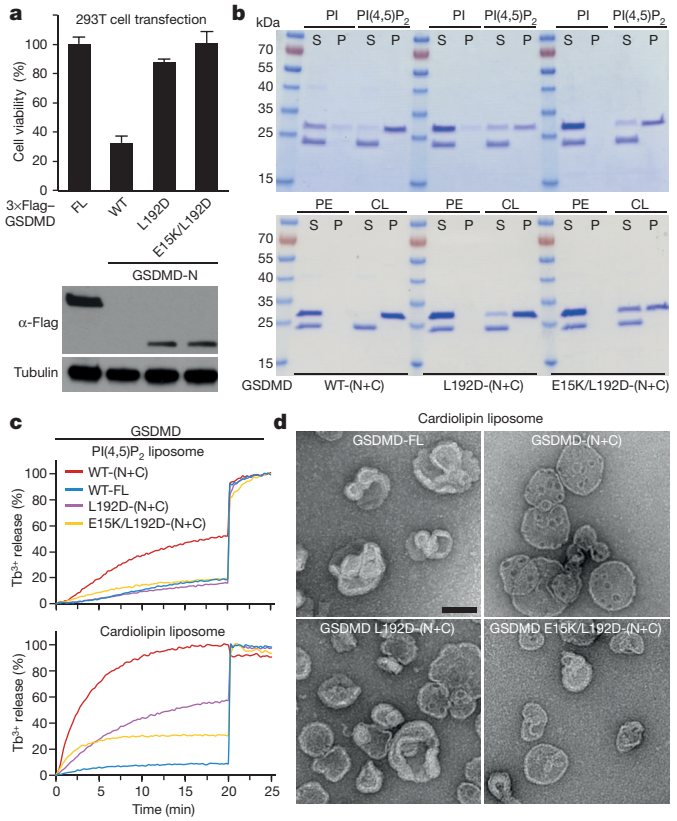


Figure 5 | Residues in the autoinhibited region in gasdermin-N are important for pyroptosis, membrane disruption and pore formation. a, Effects of L192D/E15K mutations in GSDMD-N on pyroptosis-inducing activity. Full-length GSDMD or GSDMD(1–275) (wild type or with indicated mutations) was transfected into 293T cells. ATP-based cell viability is expressed as mean ± s.d. from three technical replicates. The immunoblot shows expression of transfected GSDMD. **b**, **c**, Effects of L192D/E15K mutations on GSDMD-N lipid-binding and liposome-leakage-inducing activities. Purified GSDMD proteins were incubated with liposomes containing 80% phosphatidylcholine and 20% phosphatidylinositol, PtdIns(4,5)P_2, phosphatidylethanolamine or cardiolipin. After ultracentrifugation, the liposome-free supernatant (S) and the liposome pellet (P) were analysed by SDS–PAGE (**b**). Liposome leakage was monitored by measuring DPA chelating-induced fluorescence of released Tb^{3+} relative to that of Triton X-100 treatment (**c**). **d**, Effects of L192D/E15K mutations on pore formation by GSDMD-N. Shown are representative negative-stain electron microscopy micrographs of pores formed by indicated GSDMD proteins on cardiolipin liposomes (scale bar, 100 nm). All data shown are representative of three independent experiments.

Conserved autoinhibition in the gasdermin family

GSDMD shares about 70% homology with GSDMA3. Homology-based modelling produced a highly analogous GSDMD structure (Extended Data Fig. 7c). The structure contains a hydrophobic core resembling that in GSDMA3, in which F49/W50 play equivalent roles to F48/W49 in GSDMA3. The modelled structure bears a similar second inter-domain contact; an α4-equivalent helix has extensive hydrophobic interactions with GSDMD-C despite the residues involved being different. The residues in gasdermin-C that make the hydrophobic core are highly conserved in the gasdermin family (Extended Data Fig. 8), including above analysed L270, Y344 and A348 in GSDMA3. When equivalent residues in GSDMD (L290, Y373 and A377), GSDMA (L260, Y334 and A338), GSDMC (L319, Y398 and A402) and DFNA5 (I313, F388 and A392) were individually mutated into aspartates, 20–80% pyroptosis occurred in 293T cells expressing the mutant proteins (except for DFNA5 I313D) (Fig. 4d). Thus, structural autoinhibition is conserved in most gasdermins.

Structure–based analyses of gasdermin–N function

L184 in GSDMA3-N (L192 in GSDMD-N) on the α4 helix is contacted by the inhibitory gasdermin-C domain (Fig. 4c and Extended Data Fig. 7c). E14 in GSDMA3-N (E15 in GSDMD-N) on helix α1 is within the primary inter-domain interface (Fig. 4c). Mutations of gasdermin-C residues that contact L184 or structural regions around E14 caused constitutive activation of GSDMA3 and GSDMD[10] (Fig. 4d). As these mutations will disrupt the autoinhibition and expose L184/E14, we reasoned that L184/E14 or their adjacent residues might be important for pyroptosis. Supporting this prediction, GSDMD-N(L192D) and GSMDA3-N(L184D) induced markedly decreased pyroptosis in 293T cells (Fig. 5a and Extended Data Fig. 9a). The mutants showed evident defects in binding the liposome and causing the liposome

© 2016 Macmillan Publishers Limited. All rights reserved

leakage (Fig. 5b, c and Extended Data Fig. 9b, c). Combining the L/D mutation with the also partially inactive E15K mutation on GSDMD (E14K on GSDMA3) led to further decreased liposome binding and leakage-inducing activities. Double mutants of the GSDMD- and GSDMA3-(N+C) complexes formed fewer pores than the L/D single mutant (Fig. 5d and Extended Data Fig. 9d). GSDMD-N(L192D/E15K) and GSDMA3-N(L184D/E14K) were largely deficient in pyroptosis induction (Fig. 5a and Extended Data Fig. 9a). The two residues probably participate in oligomerization and membrane insertion during pore formation. These data reinforce the idea that the liposome-leakage and pore-forming activities of gasdermin-N are responsible for cell pyroptosis.

Discussion

We show that multiple gasdermin-N domains can induce pyroptosis owing to their pore-forming activity. Most gasdermin pores have inner diameters of 10–14 nm. The structure of GSDMA3 uncovers an autoinhibitory mechanism that is conserved in the gasdermin family. Other members of the gasdermin family may also act on endomembranes and alter other cellular physiologies. Indeed, DFNB59 and DFNA5 have been suggested to localize on the peroxisomes and mitochondria, respectively[22,23].

The gasdermin-N domain represents a new type of pore-forming protein (PFP). PFPs are diverse in sequence and present in many domains of life[24]. Most PFPs lyse cell membranes from the outside. By contrast, gasdermins and the MLKL protein (which is involved in necroptosis) kill cells from the cytosol. MLKL can induce liposome leakage, but there is no reported evidence that it can form pores[25–28]. PFPs can be divided into α-helical and β-barrel classes on the basis of the structures of their membrane-spanning regions[29]. The structure of gasdermin-N differs completely from those of α-class PFPs. A Dali search of the GSDMA3-N structure produced hits all belonging to the membrane-attack complex/perforin/cholesterol-dependent cytolysin (MACPF/CDC) family of β-class PFPs[30], but the score was not confident and no meaningful structural similarity could be identified (Extended Data Fig. 7d). Gasdermin-N is likely to use either a β-barrel-like or a distinct mechanism for pore formation, which will involve drastic conformational changes for insertion into the membrane. Our results pave the way for future studies to elucidate the structural mechanism of pore formation by gasdermins.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 March 2015; accepted 18 May 2016.
Published online 8 June 2016.

1. Jorgensen, I. & Miao, E. A. Pyroptotic cell death defends against intracellular pathogens. *Immunol. Rev.* **265,** 130–142 (2015).
2. Cookson, B. T. & Brennan, M. A. Pro-inflammatory programmed cell death. *Trends Microbiol.* **9,** 113–114 (2001).
3. Shi, J. *et al.* Inflammatory caspases are innate immune receptors for intracellular LPS. *Nature* **514,** 187–192 (2014).
4. Kayagaki, N. *et al.* Non-canonical inflammasome activation targets caspase-11. *Nature* **479,** 117–121 (2011).
5. Lamkanfi, M. & Dixit, V. M. Mechanisms and functions of inflammasomes. *Cell* **157,** 1013–1022 (2014).
6. Zhao, Y. & Shao, F. Diverse mechanisms for inflammasome sensing of cytosolic bacteria and bacterial virulence. *Curr. Opin. Microbiol.* **29,** 37–42 (2016).
7. Hagar, J. A., Powell, D. A., Aachoui, Y., Ernst, R. K. & Miao, E. A. Cytoplasmic LPS activates caspase-11: implications in TLR4-independent endotoxic shock. *Science* **341,** 1250–1253 (2013).
8. Kayagaki, N. *et al.* Noncanonical inflammasome activation by intracellular LPS independent of TLR4. *Science* **341,** 1246–1249 (2013).
9. Yang, J., Zhao, Y. & Shao, F. Non-canonical activation of inflammatory caspases by cytosolic LPS in innate immunity. *Curr. Opin. Immunol.* **32,** 78–83 (2015).
10. Shi, J. *et al.* Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* **526,** 660–665 (2015).
11. Kayagaki, N. *et al.* Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature* **526,** 666–671 (2015).
12. Saeki, N. & Sasaki, H. in *Endothelium and Epithelium: Composition, Functions, and Pathology* (eds Carrasco, J. & Matheus, M.) Ch. IX, 193–211 (Nova Science Publishers, 2011).
13. Tanaka, S., Mizushina, Y., Kato, Y., Tamura, M. & Shiroishi, T. Functional conservation of *Gsdma* cluster genes specifically duplicated in the mouse genome. *G3 (Bethesda)* **3,** 1843–1850 (2013).
14. Sato, H. *et al.* A new mutation *Rim3* resembling *Re*^{den} is mapped close to retinoic acid receptor alpha (*Rara*) gene on mouse chromosome 11. *Mamm. Genome* **9,** 20–25 (1998).
15. Porter, R. M. *et al.* Defolliculated (dfl): a dominant mouse mutation leading to poor sebaceous gland differentiation and total elimination of pelage follicles. *J. Invest. Dermatol.* **119,** 32–37 (2002).
16. Runkel, F. *et al.* The dominant alopecia phenotypes Bareskin, Rex-denuded, and Reduced Coat 2 are caused by mutations in *gasdermin 3*. *Genomics* **84,** 824–835 (2004).
17. Van Laer, L. *et al.* Nonsyndromic hearing impairment is associated with a mutation in *DFNA5*. *Nat. Genet.* **20,** 194–197 (1998).
18. Delmaghani, S. *et al.* Mutations in the gene encoding pejvakin, a newly identified protein of the afferent auditory pathway, cause DFNB59 auditory neuropathy. *Nat. Genet.* **38,** 770–778 (2006).
19. Shepard, L. A., Shatursky, O., Johnson, A. E. & Tweten, R. K. The mechanism of pore assembly for a cholesterol-dependent cytolysin: formation of a large prepore complex precedes the insertion of the transmembrane beta-hairpins. *Biochemistry* **39,** 10284–10293 (2000).
20. Fink, S. L., Bergsbaken, T. & Cookson, B. T. Anthrax lethal toxin and *Salmonella* elicit the common cell death pathway of caspase-1-dependent pyroptosis via distinct mechanisms. *Proc. Natl Acad. Sci. USA* **105,** 4312–4317 (2008).
21. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549 (2010).
22. Delmaghani, S. *et al.* Hypervulnerability to sound exposure through impaired adaptive proliferation of peroxisomes. *Cell* **163,** 894–906 (2015).
23. Van Rossom, S., Op de Beeck, K., Hristovska, V., Winderickx, J. & Van Camp, G. The deafness gene *DFNA5* induces programmed cell death through mitochondria and MAPK-related pathways. *Front. Cell. Neurosci.* **9,** 231 (2015).
24. Bischofberger, M., Iacovache, I. & van der Goot, F. G. Pathogenic pore-forming proteins: function and host response. *Cell Host Microbe* **12,** 266–275 (2012).
25. Hildebrand, J. M. *et al.* Activation of the pseudokinase MLKL unleashes the four-helix bundle domain to induce membrane localization and necroptotic cell death. *Proc. Natl Acad. Sci. USA* **111,** 15072–15077 (2014).
26. Dondelinger, Y. *et al.* MLKL compromises plasma membrane integrity by binding to phosphatidylinositol phosphates. *Cell Reports* **7,** 971–981 (2014).
27. Wang, H. *et al.* Mixed lineage kinase domain-like protein MLKL causes necrotic membrane disruption upon phosphorylation by RIP3. *Mol. Cell* **54,** 133–146 (2014).
28. Cai, Z. *et al.* Plasma membrane translocation of trimerized MLKL protein is required for TNF-induced necroptosis. *Nat. Cell Biol.* **16,** 55–65 (2014).
29. Iacovache, I., Bischofberger, M. & van der Goot, F. G. Structure and assembly of pore-forming proteins. *Curr. Opin. Struct. Biol.* **20,** 241–246 (2010).
30. Reboul, C. F., Whisstock, J. C. & Dunstone, M. A. Giant MACPF/CDC pore forming toxins: A class of their own. *Biochim. Biophys. Acta* **1858,** 475–486 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. Wei for reagents, H. Wang for suggestions on electron microscopy data analysis, and the staff of beamlines BL18U1 and BL19U1 at National Center for Protein Sciences, Shanghai, and Shanghai Synchrotron Radiation Facility for X-ray data collection. This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB08020202), the China National Science Foundation Program for Distinguished Young Scholars (31225002) and Program for International Collaborations (31461143006), and the National Basic Research Program of China 973 Program (2012CB518700 and 2014CB849602) to F.S. The research was also supported in part by an International Early Career Scientist grant from the Howard Hughes Medical Institute and the Beijing Scholar Program to F.S.

Author Contributions J.D., D.-C.W. and F.S. conceived the study; J.D., together with K.W., designed and performed the majority of the experiments; Y.S. helped with protein purification; W.L. performed the pyroptosis assay; Q.S. assisted J.D. in electron microscopy studies; J.S. provided critical reagents and suggestions; H.S. performed structural modelling; and J.D. and F.S. analysed the data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information The atomic coordinates and structure factors of GSDMA3 have been deposited in the Protein Data Bank under the accession code 5B5R. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.-C.W. (dcwang@sun5.ibp.ac.cn) or F.S. (shaofeng@nibs.ac.cn).

Reviewer Information *Nature* thanks E. Miao, F. Sigworth and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Plasmids, antibodies and reagents. Complementary DNA (cDNA) for human *GSDMA*, *GSDMB*, *GSDMC*, *GSDMD* and mouse *Gsdma3* were previously described[10]; cDNA for human *DFNA5* was obtained from Life Technologies Ultimate ORF collection (IOH41276). The cDNAs were inserted into a modified pCS2 vector with an N-terminal 3 × Flag tag or the pcDNA3 vector with a C-terminal Flag tag for transient expression in 293T cells and the pWPIlentiviral vector with an N-terminal 2 × Flag-HA tag for stable expression in iBMDM cells. For Tet-On expression, the cDNAs were inserted into a modified pLenti–NIrD vector (a gift from W. Wei) harbouring the blasticidin-resistance gene and fused with a C-terminal eGFP tag. For growth inhibition in *E. coli*, the cDNAs were cloned into the pET21a vector. For recombinant expression in *E. coli*, the cDNAs were cloned into a modified pET vector with an N-terminal 6 × His-SUMO tag or pGEX-6P-2 with an N-terminal GST tag. DNA for PFO was amplified from genomic DNA of *C. perfringens*. For recombinant expression of PFO in *E. coli*, the DNA was cloned into a pET28a vector with an N-terminal 6 × His tag. Truncation mutants of gasdermins were constructed by the standard PCR cloning strategy and inserted into the corresponding vectors with indicated tags. Point mutations were generated by the QuickChange Site-Directed Mutagenesis Kit (Stratagene). All plasmids were verified by DNA sequencing.

Antibodies used in this study include anti-Flag M2 (F4049), anti-actin (A2066) and anti-tubulin (T5168) (Sigma-Aldrich); anti-TOM20 (sc-11415) and anti-Lamp2 (sc-18822) (Santa Cruz Biotechnology); anti-Cox4 (11967, Cell Signaling Technology); anti-LAMP1 (553792) and anti-sytaxin 6 (610635) (BD Pharmingen); and anti-Na and K-ATPase α1 (2047-1, Epitomics). Rabbit antiserum for the GSDMD-C domain was generated in our in-house facility. Natural and synthetic lipid products used for liposome preparation were purchased from Avanti Polar Lipids Inc. Lipid strips used in the protein-lipid overlay assay were obtained from Echelon Biosciences Inc. Fluorescein-labelled dextran was from Life Technologies. LPS (L4524), terbium chloride ($TbCl_3$) and DPA were purchased from Sigma-Aldrich. Cell culture products were from Life Technologies and all other chemicals used were from Sigma-Aldrich unless noted.

Cell culture and transfection. HeLa and 293T cells were obtained from the American Type Culture Collection (ATCC). C57BL/6 mouse-derived wild-type and *Gsdmd*$^{-/-}$ iBMDM cells were as described[10].The cells are frequently checked by virtue of their morphological features and functionalities but have not been subjected to authentication by short tandem repeat (STR) profiling. All cell lines have been tested to be mycoplasma-negative by the commonly used PCR method. Cells were grown in DMEM supplemented with 10% (v/v) fetal bovine serum (FBS) and 2 mM L-glutamine at 37 °C in a 5% CO_2 incubator. Transient transfection in HeLa and 293T cells was performed using the JetPRIME (Polyplus Transfection) or Vigofect (Vigorous) reagents following the manufacturers' instructions. iBMDM or HeLa stable cell lines were generated by lentiviral infection as previously described[10]. Stable expression cells were sorted by flow cytometry (BD Biosciences FACSAria II) or selected by 60 μg ml^{-1} blasticidin (Invitrogen).

Microscopy imaging. To examine morphology of pyroptotic cells, cells were treated as indicated in 6-well plates (NuncProducts, Thermo Fisher Scientific Inc.). Static brightfield images of pyroptotic cells were captured using an Olympus IX71 microscope. To examine the subcellular localization of the gasdermin N-domain during pyroptosis, HeLa cells harbouring the desired Tet-On expression plasmid were treated with 2 μg ml^{-1} doxycycline in glass-bottom culture dishes (MatTek Corporation). After 3 h, live images of pyroptotic cells were recorded with a PerkinElmer UltraVIEW spinning disk confocal microscope and processed in the Volocity program. All image data shown are representative of at least three randomly selected fields.

Cell viability and osmotic protection. For viability assay, relevant cells were treated as indicated and the viability was determined by the CellTiter-Glo Luminescent Cell Viability Assay (Promega). To examine effects of osmotic protection, iBMDM cells harbouring a sensitive *Nlrp1b* allele were treated with 1.2% or 12% (w/v) osmoprotectants (PEG200, PEG1500 or PEG2000) or 40 μM zVAD for 1 h. Pyroptosis was induced by LFn-BsaK[31] or anthrax lethal toxin stimulation[32], which activate the NAIP2/NLRC4 inflammasome or NLRP1B inflammasome, respectively. Cell death was measured by the LDH release assay using the CytoTox 96 Non-Radioactive Cytotoxicity Assay kit (Promega).

Cell fractionation by differential centrifugations. Cells were collected by centrifugation at 1,000*g* for 10 min. The washed cell pellets were resuspended in 5 volumes of buffer A (20 mM HEPES, pH 7.5, 40 mMKCl, 1.5 mM $MgCl_2$, 1 mM EDTA and 250 mM sucrose) and incubated on ice for 30 min. The cells were homogenized by passing through a 22G needle 24 times. After centrifugation at 1,000*g* for 10 min, the supernatant was collected and re-centrifuged at 7,000*g* for 10 min. The supernatant and pellet were designated as the S7 and P7 fraction, respectively. The S7 fraction was centrifuged again at 20,000*g* for 20 min to obtain the S20 and P20 fractions. The S20 fraction was subjected to final centrifugation at 100,000*g* for 1 h and the supernatant was collected as the S100 fraction. The pellet was dissolved in buffer A as the P100 fraction. All centrifugations were performed at 4 °C. The fractions were solubilized in SDS loading buffer and analysed by immunoblotting as indicated.

Purification of recombinant proteins. To obtain full-length GSDMD, GSDMA and GSDMA3 proteins, *E. coli* BL21 (DE3) cells harbouring the gasdermin plasmid (pET28a-6 × His-SUMO vector) were grown in LB medium supplemented with 30 μg ml^{-1} kanamycin. Protein expression was induced overnight at 20 °C with 0.4 mM IPTG after OD_{600} reached 0.8. Cells were lysed in the buffer containing 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole and 10 mM 2-mercaptoethanol. The fusion protein was affinity-purified by Ni-Sepharose beads (GE Healthcare Life Sciences). The SUMO tag was removed by overnight digestion with homemade ULP1 protease at 4 °C. The untagged protein was further purified by HiTrap Q anion exchange and Superdex G75 gel filtration chromatography (GE Healthcare Life Sciences). Selenomethionine-substituted (SeMet) GSDMA3 was expressed in the methionine auxotrophic *E. coli* strain B834 (DE3) and purified in the same way as the native protein.

The engineered gasdermin proteins (GSDMD, GSDMA and GSDMA3) containing the PreScission protease (PPase) recognition site were expressed and purified by following the same procedure as that for native gasdermin proteins. Inter-domain cleavage was performed by overnight digestion with homemade PPase at 4 °C. The proteins were further purified by Superdex G75 gel-filtration chromatography to obtain high-quality noncovalent complex of the cleaved gasdermin. To obtain the GSDMA-N domain alone, the noncovalent complex of GSDMA was further subjected to HiTrap Q anion exchange chromatography, followed by another round of Superdex G75 gel filtration chromatography. To obtain the gasdermin-C domain of GSDMA3 and GSDMA, the flow-through fractions of PPase-cleaved GSDMA3 and GSDMA proteins from Ni-Sepharose beads were subjected to HiTrap Q anion exchange and Superdex G75 gel filtration chromatography. To obtain GSDMD-C protein, *E. coli* BL21 (DE3) cells were transformed with pGEX-6P-2-GSDMD (residues 276–484). The GST-tagged protein was purified by affinity chromatography using glutathione-Sepharose beads (GE Healthcare Life Sciences) and the tag was removed by overnight digestion with PPase at 4 °C. The proteins were further purified by Superdex G75 gel filtration chromatography. Recombinant PFO was expressed and purified by following the same procedure as that for the gasdermin protein. All the purified proteins were concentrated and stored in the buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl and 5 mM dithiothreitol.

Bacterial growth inhibition and protoplast lysis. To assay the cytotoxicity of the gasdermin-N domain in *E. coli*, equal amounts of *E. coli* BL21 (DE3) cells were transformed with 0.1 μg of indicated plasmid. The transformed cells were serially diluted and plated onto LB agar containing the appropriate antibiotics in the presence or absence of IPTG. The colony-forming unit (CFU) was determined by counting the number of viable bacteria per transformation after overnight culture at 37 °C.

To prepare the protoplast, *B. megaterium* cells were grown in AB3 medium (DIFCO) at 37 °C until the OD_{600} reached 2.0. The centrifuged bacterial pellets, resuspended in the buffer containing 20 mM sodium malate (pH 6.5), 20 mM $MgCl_2$ and 500 mM sucrose, were incubated with 2 mg ml^{-1} lysozyme at 37 °C until protoplast formation was complete (judged under phase-contrast microscopy).The protoplasts were diluted to an OD_{600} of 1.0. To assay protoplast lysis, aliquots of the protoplasts (1.9 ml) were incubated with 100 μl of indicated gasdermin proteins (final concentration of 0.6 μM) at 37 °C for 30 min. To achieve complete lysis of the protoplast, 100 μl of 2% (v/v) Triton X-100 was added. The OD_{600} of each protoplast aliquot before and after incubation with the gasdermin protein was measured and defined as $A_0$and A_n, respectively, and that after Triton X-100 treatment was treated as A_{100}. The percentage of protoplast lysis is defined as: lysis (%) = $(A_0 - A_n) \times 100/(A_0 - A_{100})$. For the time-course plot of protoplast lysis, the OD_{600} of each protoplast aliquot was continuously recorded for 15 min at 1-min intervals before Triton X-100 was added.

Liposome preparation. Phospholipids and phosphoinositides were dissolved in chloroform and chloroform-methanol mixture (20:9, v-v), respectively. Lipids with indicated compositions (0.5 μmol) were mixed in a glass vial. The solvent was evaporated under a stream of nitrogen, and the dry lipid film was then hydrated at room temperature with constant mixing in 500 μl buffer 1 (20 mM HEPES (pH 7.5) and 150 mM NaCl). Liposomes were generated by extrusion of the hydrated lipids through a 100-nm polycarbonate filter (Whatman) 35 times using the Mini-Extruder device (Avanti Polar Lipids Inc.). For Tb^{3+}-encapsulated liposomes, the lipid film was hydrated with 500 μl buffer 2 (20 mM HEPES (pH 7.5), 100 mM NaCl, 50 mM sodium citrate and 15 mM $TbCl_3$). After the extrusion

© 2016 Macmillan Publishers Limited. All rights reserved

process, Tb^{3+} ions outside the liposome were removed by repeated washing with buffer 2 on a centrifugal filter device (Amicon Ultra-4, 100K MWCO, Millipore). The liposomes were subjected to buffer change into buffer 1 for use. To prepare dextran-encapsulated liposomes, the lipid film was hydrated in buffer 1 supplemented with 2 mg ml^{-1} dextrans. The liposomes were repeatedly washed to remove external dextran and then resuspended in 500 μl buffer 1. All the liposomes were stored at 4 °C and used within 48 h.

Liposome-binding and leakage assays. The indicated gasdermin proteins (5 μM) were incubated with the indicated liposome (500 μM lipids) at room temperature for 30 min in a total volume of 80 μl. Samples were centrifuged in a Beckman Optima MAX-XP ultracentrifuge at 4 °C for 20 min at 100,000*g*. The supernatant (S) was collected to examine proteins not bound to the liposome. The pellets (P) were washed twice with 100 μl buffer 1 by re-centrifugation and brought up to the same volume as the supernatant. The S and P fractions were analysed by SDS–PAGE followed by Coomassie blue staining.

For liposome leakage assay, aliquots of Tb^{3+}-encapsulated liposomes were diluted to 300 μM lipid concentration in 90 μl of buffer 1 supplemented with 15 μM DPA. The excitation and emission wavelengths of 270 nm and 490 nm, respectively, were used to examine the Tb^{3+}/DPA chelates[33]. The emission fluorescence before adding the gasdermin protein was treated as F_{t0}. 10 μl of protein was then added to a final concentration of 0.6 μM, and the emission fluorescence was continuously recorded as F_t at 15-s intervals. After 20 min, 10 μl of 1% Triton X-100 was added to achieve complete release of Tb^{3+}, and mean values of the top three fluorescence reads were defined as F_{t100}. The percentage of liposome leakage at each time point is defined as: leakage (t) (%) $= (F_t - F_{t0}) \times 100/(F_{t100} - F_{t0})$. For dextran leakage assay, aliquots of dextran-encapsulated liposomes (300 μM of lipids) were incubated with 0.6 μM indicated proteins at room temperature for 30 min in a total volume of 100 μl. After centrifugation, the released dextran in the supernatants was collected and the emission fluorescence (521 nm) after excitation at 494 nm was measured as F_n. The emission fluorescence of supernatants of untreated liposomes was measured as F_0, and that of the liposomes solubilized with 0.1% Triton X-100 was defined as F_{100}. The percentage of dextran leakage is defined as: leakage (%) $= (F_n - F_0) \times 100/(F_{100} - F_0)$.

Crosslinking assays of gasdermin-N oligomerization. To assay gasdermin-N domain oligomerization *in vitro*, indicated PPase-cleaved engineered gasdermin proteins, before or after incubation with the liposome, were treated with 5 mM glutaraldehyde for 30 min at room temperature. The liposome pellets were solubilized in SDS loading buffer. Samples with or without crosslinking were analysed by SDS-agarose gel electrophoresis as previously described[19]. To assay the oligomerization during pyroptosis, relevant cells, before or after pyroptosis induction, were harvested in PBS. The cell pellets were homogenized by passing through a 22G needle 24 times. Cell lysates were centrifuged at 100,000*g* for 1 h to obtain the supernatant and the pellet fractions. The pellet was homogenized by gentle sonication. Both fractions were treated with 2 mM glutaraldehyde at room temperature for 15 min. Samples with or without crosslinking were analysed by both SDS-agarose and SDS–PAGE gel electrophoresis followed by immunoblotting.

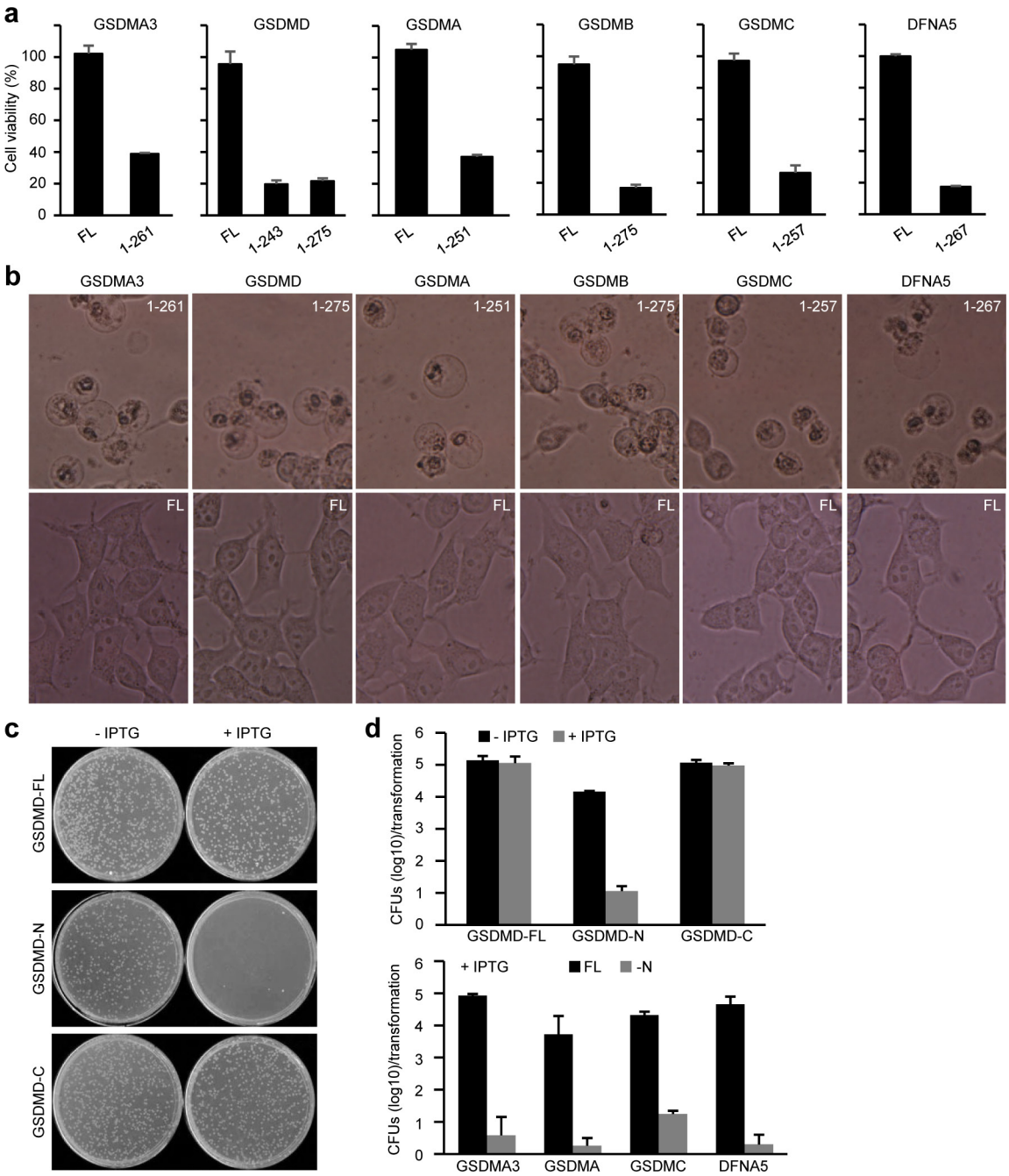
Electron microscopy and image processing. Gasdermin proteins (5 μM) were incubated with indicated liposomes (500 μM of lipids) at room temperature for 30 min. Aliquots of the mixture (5 μl) were transferred to carbon support films on electron microscopy grids and negatively stained with 2% uranyl acetate. Samples were imaged on a Tecnai T12 microscope (FEI) at 120 kV. Images were taken on a Gatan 4k × 4k CCD camera with a nominal magnification of 30,000×, giving a final pixel size of 3.71 Å. To prepare pores on lipid monolayers, solutions of noncovalent complexes of cleaved GSDMD (500 nM) and GSDMA3 (100 nM) were pipetted into Teflon wells and coated with a droplet of 1 mM lipid mixture containing 80% phosphatidylcholine and 20% cardiolipin in chloroform. Negative-stain electron microscopy of the GSDMA3 pores was performed and images were captured as above described. Complete and undistorted pore particles were manually selected from the micrographs using EMAN2 (ref. 34). A total of 7,056 particle images were collected and normalized. After determining the defocus, the particles were phase-flipped for contrast transfer function correction using EMAN2 (ref. 34). Two-dimensional reference-free classification was then performed in Relion1.3 (ref. 35). More than 80% of class averages were pores of a uniform size around 30 nm in diameter. The averaged view with the best particle contrast of a class comprising 242 particles was selected and its rotational auto-correlations were calculated in SPIDER[36] to determine the symmetry.

Structure determination. The crystallization experiments were performed using the sitting-drop vapour diffusion method at 20 °C with 2-μl drops containing 1 μl protein solution and 1 μl reservoir solution equilibrated over 100 μl reservoir solution. Initial crystallization hits of GSDMA3 were found from the Index Kit (Hampton Research) screen. Qualified crystals of SeMet-labelled GSDMA3 were obtained in the reservoir buffer containing 100 mMBis-Tris (pH 6.5), 19% polyethylene glycol 3550, and 10 mM TCEP within 2 weeks. For data collection, the crystals were soaked in a cryoprotectant solution containing the reservoir buffer supplemented with 20% ethylene glycol before flash-freezing with liquid nitrogen. Diffraction data were collected at the Shanghai Synchrotron Radiation Facility (Shanghai, China) beamline BL19U1 under a wavelength of 0.97776 Å, and processed with the HKL 2000 suite[37]. Phase determination by the single wavelength anomalous dispersion (SAD) method and automatic model building were performed in PHENIX[38]. The rest of the model was manually built with Coot[39]. The structure of GSDMA3 was refined in PHENIX, and manual modelling was performed between refinement cycles. The statistics of data collection and refinement are summarized in Extended Data Table 1.The quality of the final model was validated by MolProbity[40]. Ramachandran statistics indicated that all the residues are in the allowed region, in which 97.25% fell into the favoured region. Each asymmetric unit contains one GSDMA3 molecule and the model covers residues 1–453 of the 464 total residues. As well as the C-terminal tail being absent, four loops (residues 66–80, 170–178, 188–193 and 249–263) could not be modelled owing to the lack of interpretable electron density for these highly flexible loops.

Homology-based modelling of GSDMD structure was performed with the program Modeller[41] based on the sequence alignment of GSDMA3 and GSDMD. The structural model of GSDMA3 was completed by using Loop Refine in Modeller and used as the modelling template. The top hit of GSDMD models was evaluated using the DOPE statistical potential score[42] and chosen for subsequent analysis and comparison with GSDMA3 structure. All structural figures were generated in PyMOL (http://www.pymol.org).

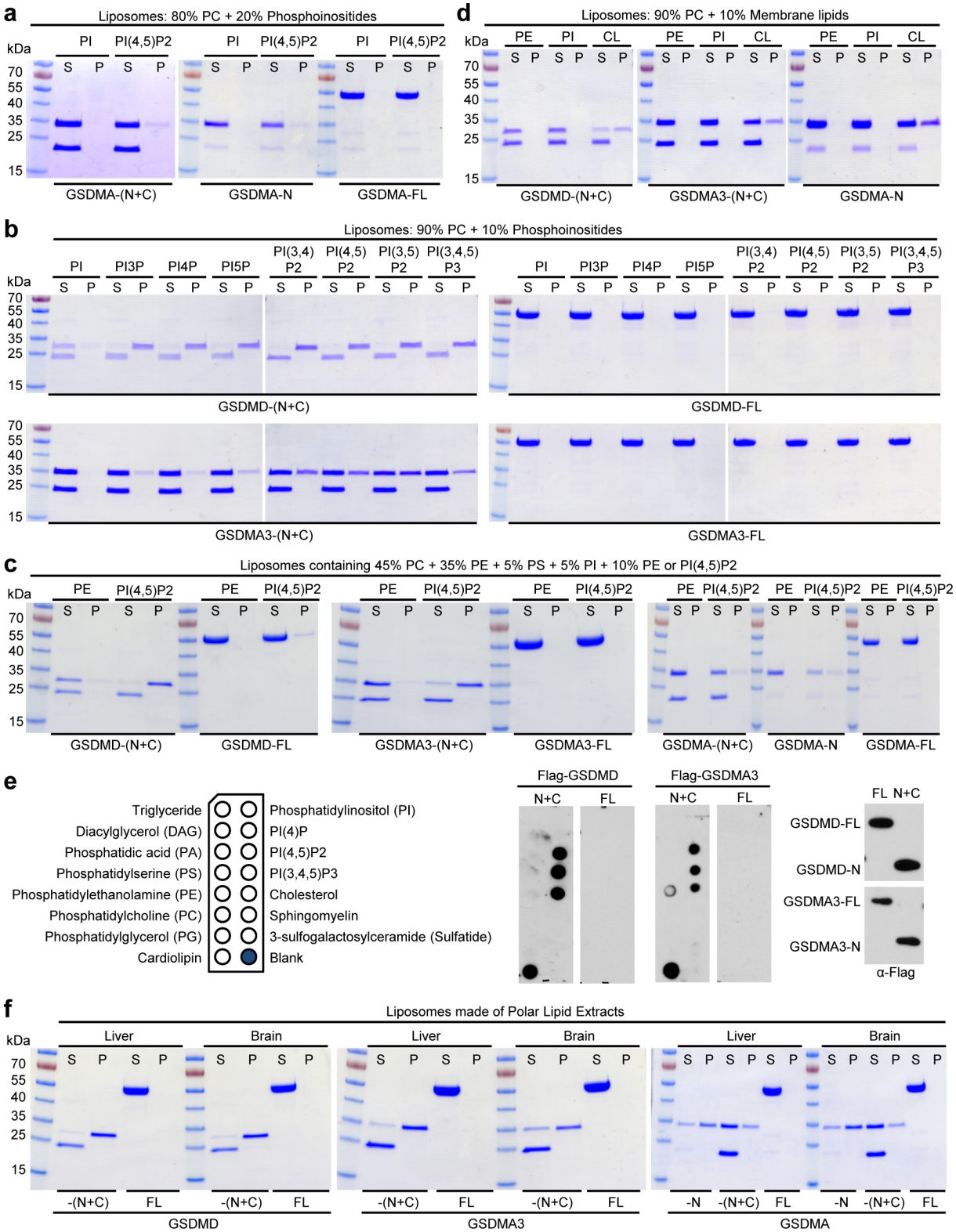
31. Zhao, Y. *et al.* The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477,** 596–600 (2011).
32. Gong, Y.-N. *et al.* Chemical probing reveals insights into the signaling mechanism of inflammasome activation. *Cell Res.* **20,** 1289–1305 (2010).
33. Wilschut, J. & Papahadjopoulos, D. Ca^{2+}-induced fusion of phospholipid vesicles monitored by mixing of aqueous contents. *Nature* **281,** 690–692 (1979).
34. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
35. Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415,** 406–418 (2012).
36. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
37. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (eds Carter, C. W. Jr & Sweet, R. M.) 307–326 (Academic, 1997).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66,** 213–221 (2010).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
40. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66,** 12–21 (2010).
41. Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Curr. Protoc. Protein Sci.* Chapter 2, Unit 2.9 (2007).
42. Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15,** 2507–2524 (2006).

© 2016 Macmillan Publishers Limited. All rights reserved



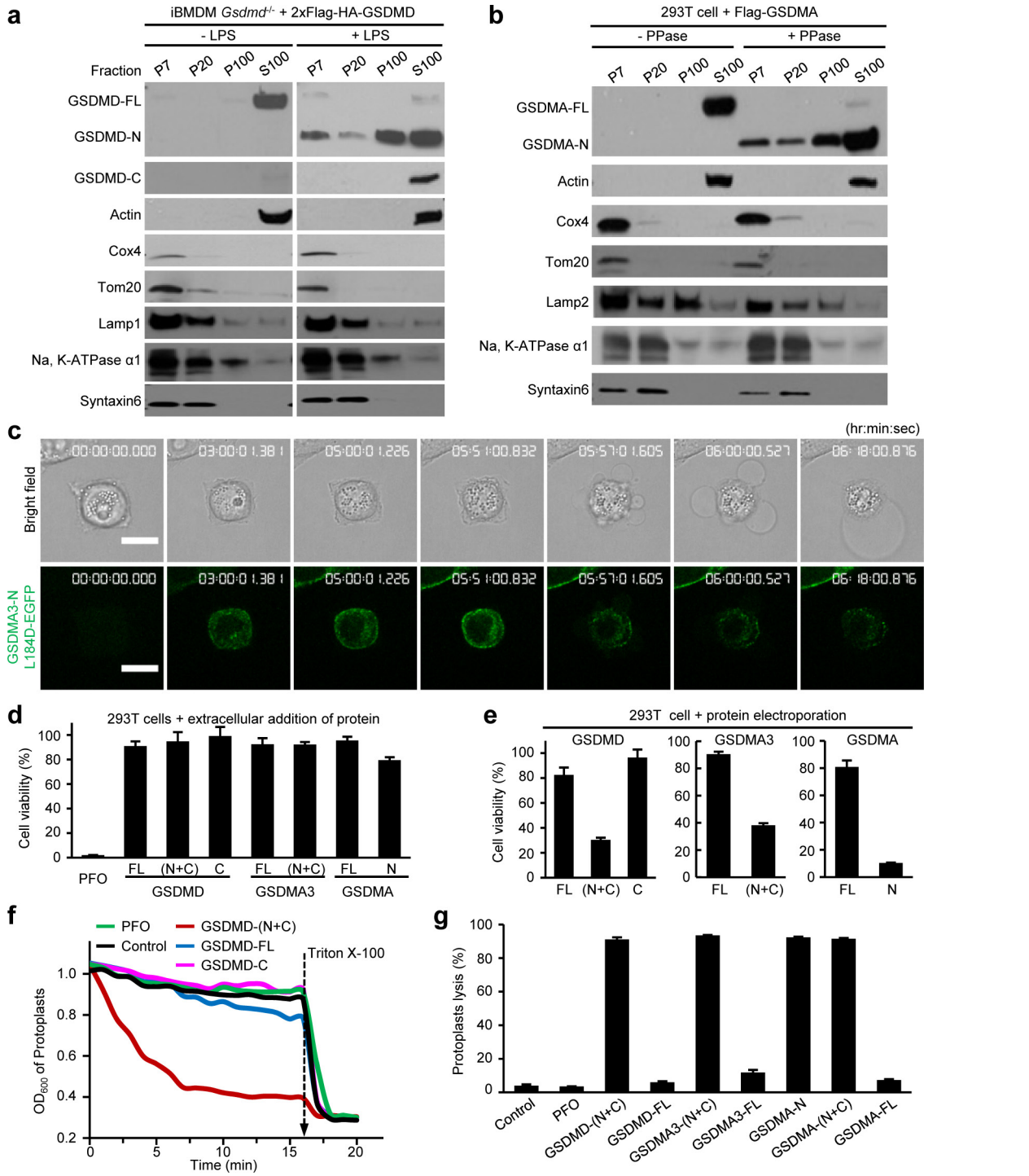
Extended Data Figure 1 | Multiple gasdermin-N domains can induce mammalian cell pyroptosis and also exhibit cytotoxicity in bacteria. **a**, **b**, Full-length (FL) or N-terminal domain regions of different gasdermin-family members were transfected into 293T cells for 20 h. Human GSDMD and mouse GSDMA3 had an N-terminal 3 × Flag tag and human GSDMA, GSDMB, GSDMC and DFNA5 had a C-terminal Flag tag. ATP-based cell viability is expressed as mean ± s.d. from three technical replicates (**a**). Representative views of cell death morphology are shown in **b**. **c**, **d**, Cytotoxicity of the gasdermin-N domain in bacteria. Indicated gasdermins were cloned into an IPTG-inducible vector for transformation into *E. coli*. **c**, Representative agar plates showing transformed *E. coli* colonies for GSDMD. **d**, Bacterial colony-forming units (CFU) per transformation for GSDMD and other gasdermins are shown in the logarithmic form ($\log_{10}$) as mean ± s.d. from three technical replicates. All data shown are representative of three independent experiments.

© 2016 Macmillan Publishers Limited. All rights reserved



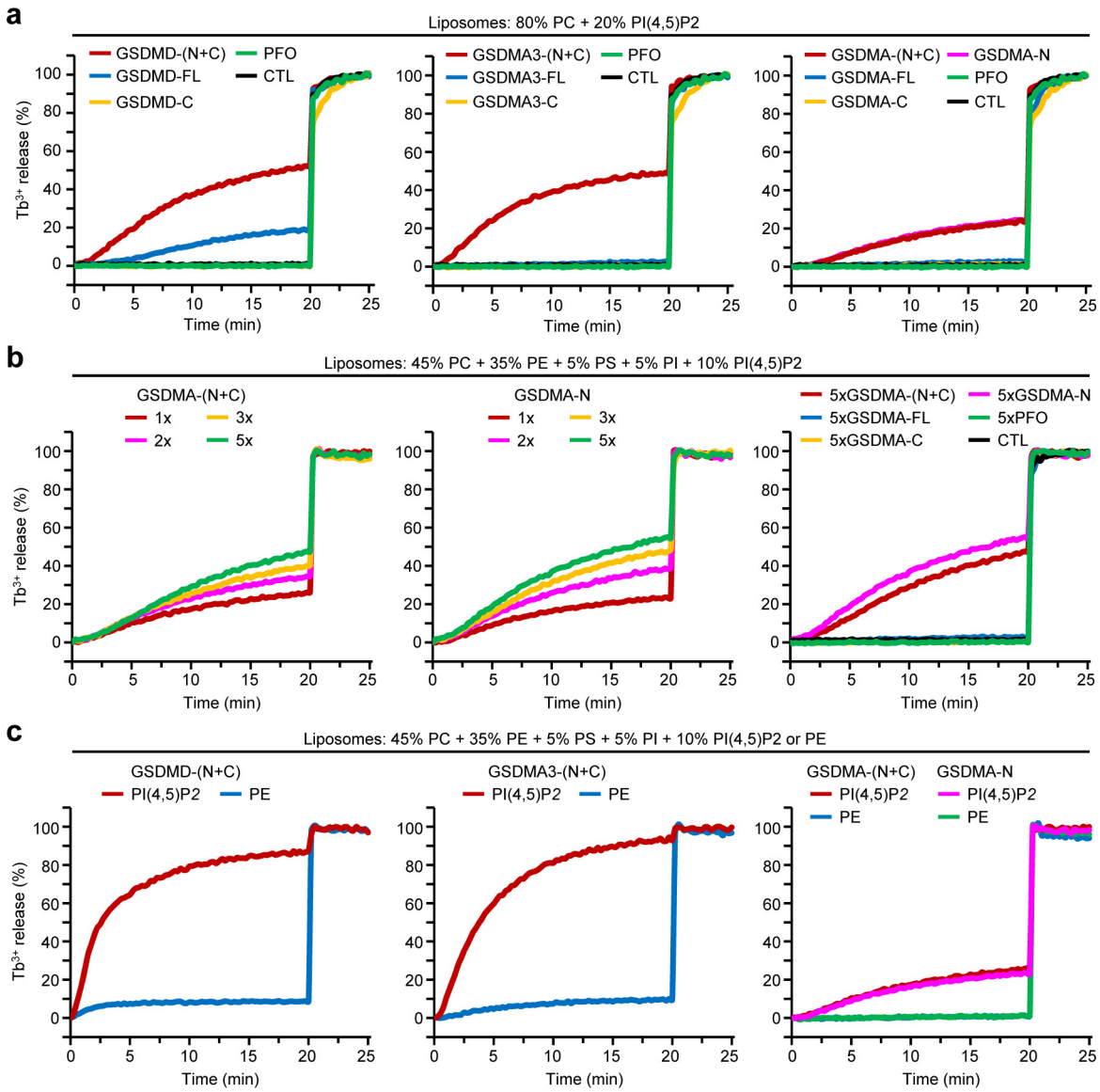
Extended Data Figure 2 | Membrane phospholipid binding of the gasdermin-N domain. a–d, f, Liposomes with indicated lipid compositions (**a–d**) or prepared using bovine liver or brain-derived polar lipid extracts (**f**) were incubated with purified gasdermin proteins. After ultracentrifugation, the liposome-free supernatant (S) and liposome pellet (P) were analysed by SDS–PAGE and Coomassie blue staining. **e**, Noncovalent complex of cleaved GSDMD and GSDMA3 with a Flag tag attached to the end of the gasdermin-N domain or the corresponding uncleaved full-length proteins were incubated with the lipid strips, and the strips were then probed with the anti-Flag antibody. Right, protein loading control. All data shown are representative of three independent experiments.

© 2016 Macmillan Publishers Limited. All rights reserved



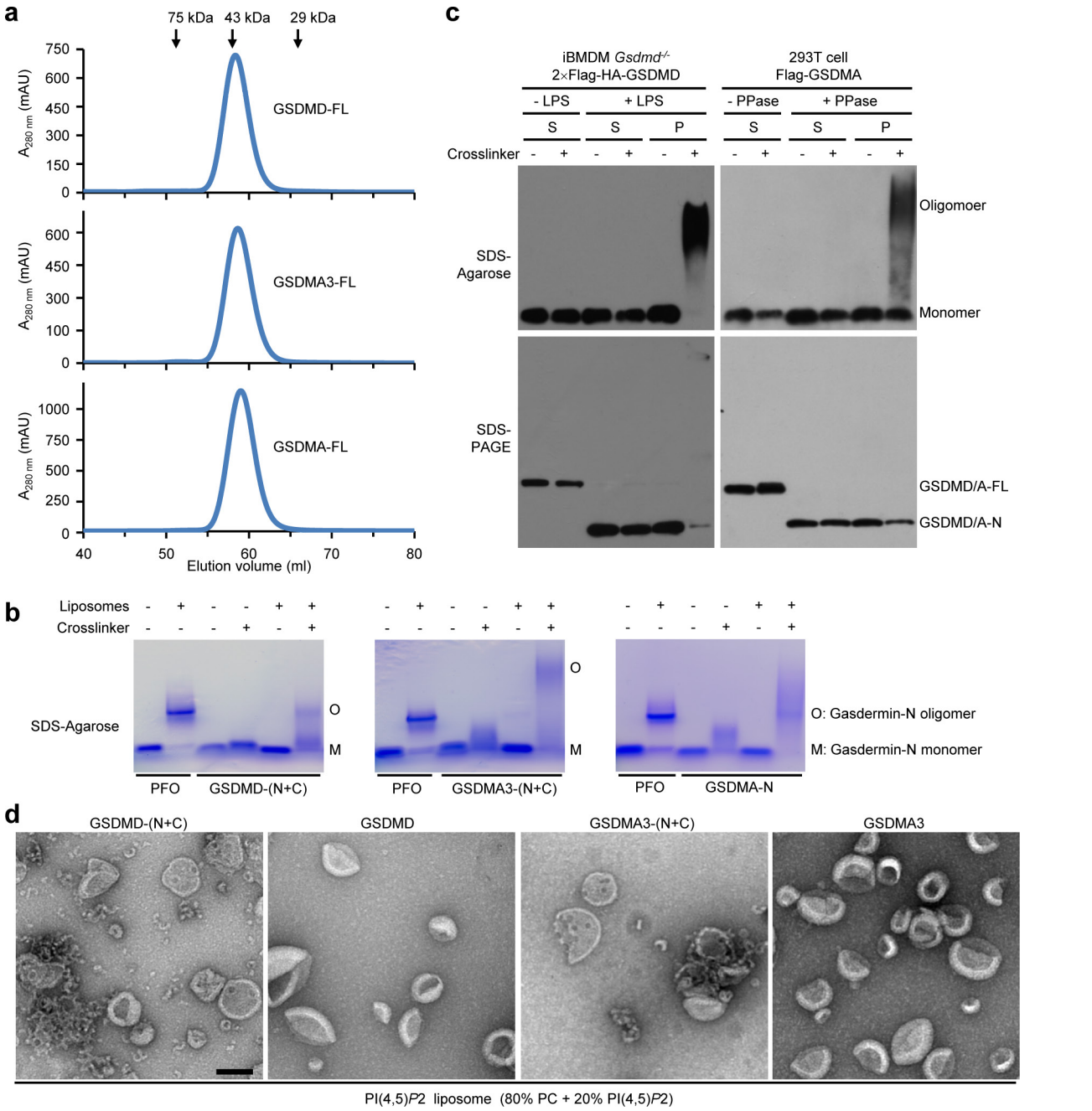
Extended Data Figure 3 | Biomembrane association and lysis by the gasdermin-N domain. a, **b**, Subcellular fractionation of the gasdermin-N domains of GSDMD and GSDMA during pyroptosis. *Gsdmd*$^{-/-}$ iBMDMs expressing 2 × Flag and haemagglutinin (HA)-tagged GSDMD were untreated or stimulated with LPS electroporation (**a**). 293T cells expressing PPase-cleavable Flag–GSDMA were untreated or electroporated with purified PPase (**b**). Homogenized cell extracts were sequentially centrifuged at 700*g*, 20,000*g* and 100,000*g* to separate membrane fractions (P7, P20 and P100) from the S100 soluble fraction. The factions were immunoblotted as indicated. **c**, Microscopy of GSDMA3-N domain localization in cells undergoing pyroptosis. The gasdermin-N domain of GSDMA3 (GSDMA3-N(L184D)) fused N-terminally to eGFP was stably expressed in HeLa cells under a tetracycline-inducible promoter. Shown are representative time-lapse cell images (brightfield and fluorescence) taken from 4–5 h after doxycycline addition. Scale bar, 15 μm. For videos of two representative cells, see Supplementary Videos 3 and 4. **d**, **e**, Effects of extracellular or intracellular delivery of purified gasdermin proteins on 293T cell viability. Equal amounts of indicated gasdermin proteins or PFO were added directly into cell culture medium (**d**) or electroporated into the cytosol (**e**). ATP-based cell viability is expressed as mean ± s.d. from three technical replicates. **f**, **g**, Bacterial protoplast lysis by purified gasdermin proteins. Protoplasts of *B. megaterium* were treated with indicated gasdermin proteins or PFO. Membrane lysis was assessed by measuring the OD_{600} of the protoplasts. Triton-X 100 treatment was used to achieve 100% lysis of the protoplasts. Time-course measurement of GSDMD treatment is shown in **f**. Relative protoplast lysis by GSDMD and other gasdermins is expressed as mean ± s.d. from three technical replicates (**g**). All data shown are representative of three independent experiments.

© 2016 Macmillan Publishers Limited. All rights reserved



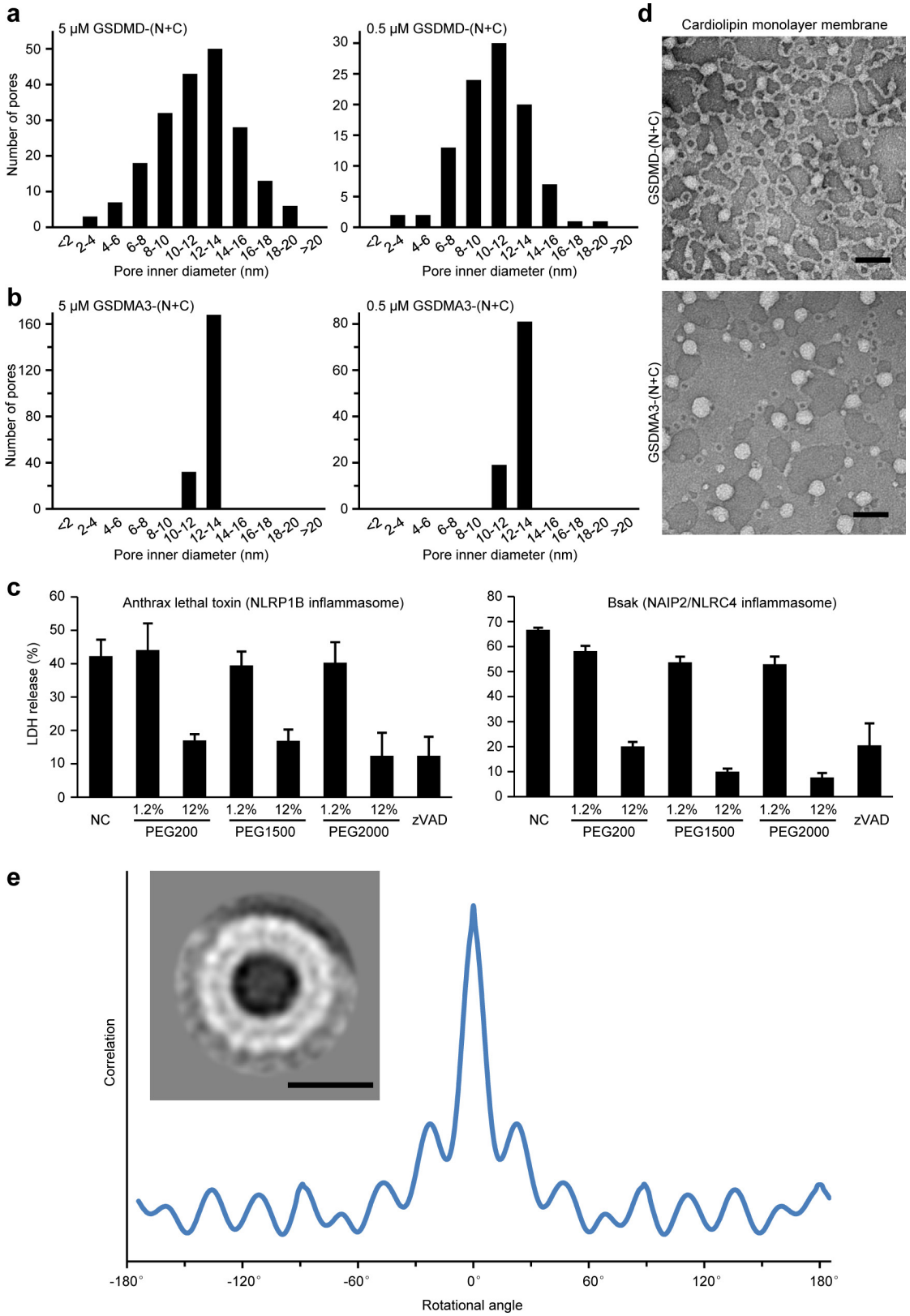
Extended Data Figure 4 | Liposome-leakage-inducing activity of the gasdermin-N domain. a–c, Liposomes with indicated lipid compositions were treated with purified gasdermin proteins or PFO as indicated. Liposome leakage was monitored by measuring DPA chelating-induced fluorescence of released Tb^{3+}. Time course of relative Tb^{3+} release is shown. A dose titration of GSDMA proteins is shown in **b**. CTL, control. All data shown are representative of three independent experiments.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | Membrane binding-induced oligomerization of and pore formation by the gasdermin-N domain. a, Gel filtration chromatography of full-length GSDMD, GSDMA and GSDMA3. Indicated gasdermin proteins were loaded on the Superdex G75 column. Arrows indicate elution volume of the molecular mass markers. **b**, Oligomerization of gasdermin-N domain on the liposome membrane. Indicated gasdermin proteins or PFO were incubated with cardiolipin or cholesterol liposomes, respectively. Intact proteins or proteins associated with the liposomes were mock treated or treated with glutaraldehyde and analysed by SDS-agarose gel electrophoresis and Coomassie blue staining. The gasdermin-C domain migrating at the bottom of the gel was omitted for clarity. **c**, Oligomerization of the gasdermin-N domain during pyroptosis. To trigger pyroptosis, *Gsdmd*$^{-/-}$ iBMDMs expressing 2 × Flag-HA–GSDMD and HeLa cells expressing the PPase-cleavable Flag–GSDMA were electroporated with LPS and PPase, respectively. The cytosol (S) and membrane (P) fractions from unstimulated and pyroptotic cells were subjected to glutaraldehyde-mediated crosslinking followed by SDS-agarose (top) or SDS–PAGE (bottom) gel electrophoresis. **d**, Pore-forming activity of the gasdermin-N domain. Liposomes with 80% phosphatidylcholine and 20% PtdIns(4, 5)P_2 were treated with indicated gasdermin proteins. Shown are representative negative-stain electron microscopy micrographs of the liposomes (scale bar, 100 nm). All data shown are representative of three independent experiments.

© 2016 Macmillan Publishers Limited. All rights reserved

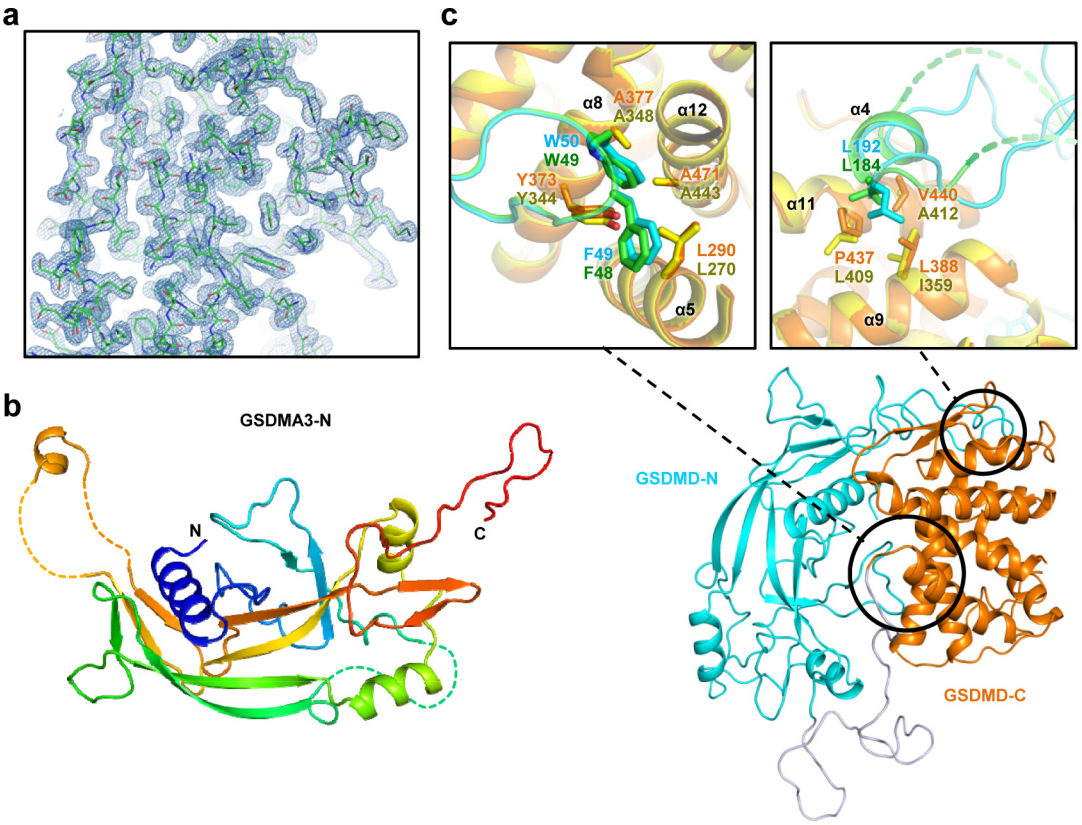


Extended Data Figure 6 | See next page for caption.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Figure 6 | Analyses of the gasdermin pore. a, **b**, Size distribution of GSDMD and GSDMA3 pores formed on cardiolipin liposomes. The inner diameters of pores were measured and plotted. A total of 200 or 100 pores for 5 μM or 0.5 μM GSDMD/GSDMA3-treated liposome samples, respectively, were randomly selected from the negative-stain electron microscopy micrographs in Fig. 3a. **c**, Effects of different PEG molecules on lactate dehydrogenase (LDH) release from caspase-1-mediated pyroptotic cells. iBMDMs harbouring a sensitive *Nlrp1b* allele were treated with indicated mass concentration of different PEG molecules and then stimulated with anthrax lethal toxin or LFn-BsaK to activate the canonical NLRP1B or NAIP2/NLRC4 inflammasomes, respectively. 1.2% PEG200 and 12% PEG2000 (mass concentration) have roughly the same molar concentration. Shown are LDH release expressed as mean ± s.d. from three technical replicates. **d**, Pores formed by active GSDMD and GSDMA3 on monolayer membranes containing 80% phosphatidylcholine and 20% cardiolipin. Shown are representative negative-stain electron microscopy micrograph images (scale bar, 100 nm). **e**, Symmetry determination of the gasdermin pore. GSDMA3 pores formed on the monolayer membrane (**d**) were subjected to 2D reference-free classification. One class of pores with the best particle contrast were subjected to rotational auto-correlation calculation and the inlet electron microscopy image (scale bar, 20 nm) shows the averaged view of the class of pores (242 particles). The analyses revealed 16-fold symmetry. Data shown in **a**–**d** are representative of three independent experiments.

© 2016 Macmillan Publishers Limited. All rights reserved

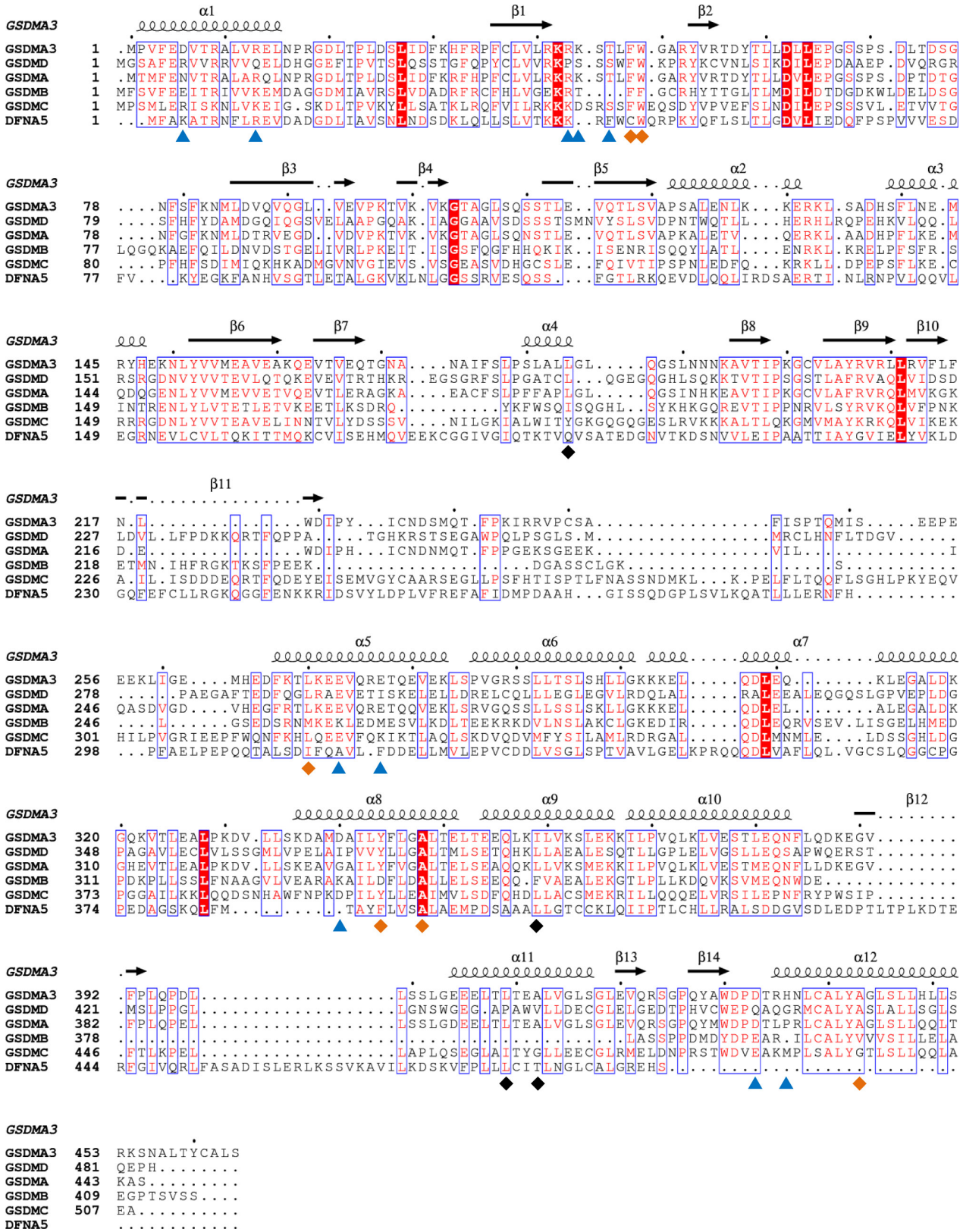


d

No.	PDB chain	Z score	RMSD (Å)	Aligned residues	Identity (%)	Molecule name
1	3hvn-A	5.0	5.4	130	11	Suilysin
2	4cdb-A	4.9	5.4	127	11	Listeriolysin O
3	4qqa-A	4.9	5.9	130	15	Pneumolysin
4	3cqf-A	4.7	5.9	127	10	Anthrolysin O
5	1m3j-A	4.4	5.3	129	12	Perfringolysin O
6	1s3r-A	4.3	5.6	129	11	Intermedilysin
7	3kk7-A	3.4	5.2	116	9	MACPF from *B. thetaiotaomicron*
8	2qp2-A	3.3	5.4	117	11	MACPFfrom *P. luminescens*
9	4wvm-B	3.2	3.4	84	4	Stonustoxin subunit beta
10	4wvm-A	3.1	3.2	85	11	Stonustoxin subunit alpha
11	2qqh-A	3.1	5.1	115	5	Complement component C8 alpha chain
12	3nsj-A	3.0	4.9	111	8	Perforin-1
13	3ojy-B	2.9	5.2	118	8	Complement component C8 beta chain
14	4ov8-A	2.8	6.6	110	4	Pleurotolysin B
15	4e0s-B	2.7	4.9	114	11	Complement component C6

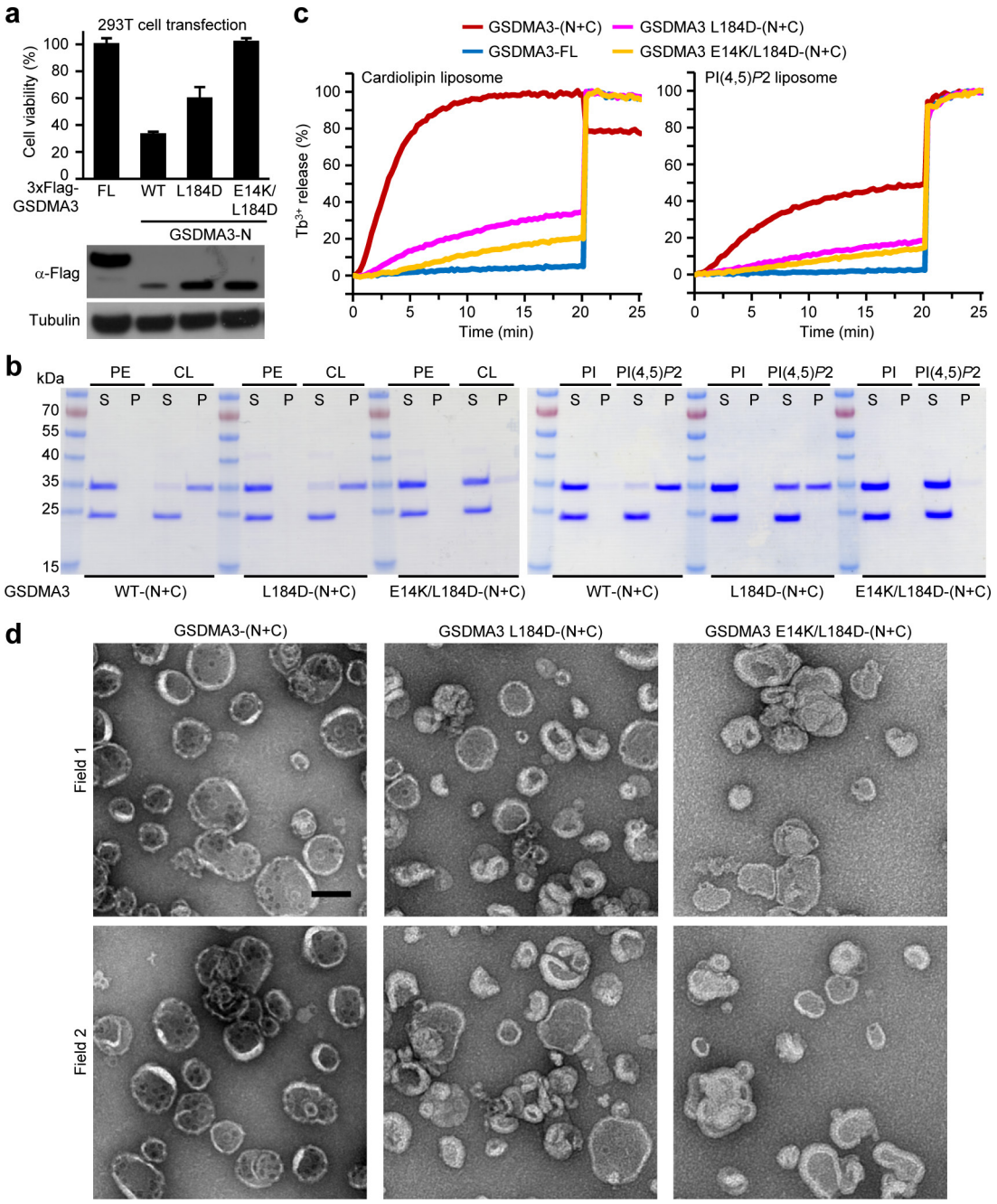
Extended Data Figure 7 | Crystal structure of GSDMA3 and Dali search results for its gasdermin-N domain. a, $2F_o - F_c$ electron density map (contoured at 1.0σ) of GSDMA3 gasdermin-N domain (GSDMA3-N) structure. **b**, Cartoon diagram of GSDMA3-N structure. **c**, Structural model of GSDMD obtained from homology modelling and the conserved autoinhibitory interactions. Bottom, overall structure of modelled GSDMD; top, comparisons of the hydrophobic core (left) and the second inter-domain contact (right) with the corresponding structures in GSDMA3. Conserved residues involved in the autoinhibitory interactions are labelled and shown as sticks. Cyan, GSDMD-N; orange, GSDMD-C; green, GSDMA3-N; yellow, GSDMA3-C. **d**, Dali search results for the GSDMA3-N structure.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | Multiple sequence alignment of gasdermin family members. GSDMA3 is a mouse protein and sequences of other gasdermins are from human. The secondary structures determined from GSDMA3 are marked along the sequence. The alignment was performed with the ClustalW2 algorithm with structure-based manual adjustment of the α4 region in GSDMA3 and GSDMD. Identical residues are highlighted by dark red background and conserved residues are coloured red. The residues involved in the autoinhibitory interactions are marked underneath the sequences with blue triangle for polar residues, orange rhombus for hydrophobic residues and black rhombus for hydrophobic residues in the second inter-domain interface. The residue number is indicated on the left of the sequence.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Mutations in GSDMA3-N affecting lipid binding and pore formation also reduce pyroptosis. a, Effects of L184D/E14K mutations on pyroptosis-inducing activity of GSDMA3-N (residues 1–284). Full-length GSDMA3 or its gasdermin-N domain (wild type or indicated mutants) was transfected into 293T cells. ATP-based cell viability is expressed as mean ± s.d. from three technical replicates. **b**, **c**, Effects of L184D/E14K mutations on lipid-binding and liposome-leakage-inducing activities of GSDMA3-N domain. Liposomes containing 80% phosphatidylcholine and 20% phosphatidylethanolamine, cardiolipin, phosphatidylinositol or PtdIns(4,5)P_2 were treated with purified GSDMA3. After ultracentrifugation, the liposome-free supernatant (S) and the liposome pellet (P) were analysed by SDS–PAGE (**b**). Liposome leakage was monitored by measuring DPA chelating-induced fluorescence of released Tb^{3+} (**c**). Triton-X 100 treatment was used to achieve 100% leakage. **d**, Effects of L184D/E14K mutations on pore formation by GSDMA3-N. Representative electron microscopy images of the pores on the cardiolipin liposome are shown (scale bar, 100 nm). All data shown are representative of three independent experiments.

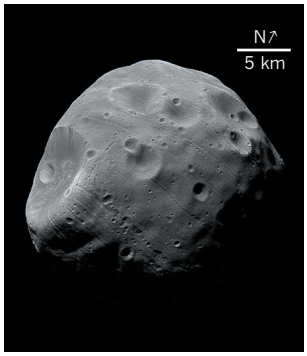
© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Data collection refinement statistics

	Se_GSDMA3(1-464) (5B5R)
Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	43.453, 103.414, 49.625
α, β, γ (°)	90.00, 110.57, 90.00
Wavelength (Å)	0.97776
Resolution (Å)[a]	50.00-1.90 (1.93-1.90)[a]
R_{merge}	0.071 (0.876)
$I/\sigma(I)$	25.6 (2.4)
Completeness (%)	97.1 (92.7)
Redundancy	6.6 (5.1)
Refinement	
Resolution (Å)	37.91-1.90
No. reflections	44,110
R_{work} / R_{free}	0.1892/0.2319
No. atoms	
Protein	3,264
Water	218
B factors	
Protein	28.36
Water	33.91
r.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.848

One SeMet crystal was used for data collection and structure determination.
*Values in parentheses are for highest-resolution shell.

© 2016 Macmillan Publishers Limited. All rights reserved



ESA/DLR/FU BERLIN (G. NEUKUM)

PLANETARY SCIENCE

Martian moons formed *in situ*

The moons of Mars may have formed from a disk of debris kicked up by the impact of a giant meteorite on the planet.

Astronomers have struggled to explain the existence of Phobos (**pictured**) and Deimos, the small, irregularly shaped moons of the red planet. One view is that they were asteroids captured by Mars. But a team led by Pascal Rosenblatt at the Royal Observatory of Belgium in Brussels tested an alternative idea using computer simulations of how orbiting debris, created by a giant impact, might coalesce.

Relatively large moons form in the inner part of the disk thrown up by such a smash, and migrate outward, causing the outer part of the disk to coalesce into two bodies the sizes of Phobos and Deimos. The inner large moons are eventually dragged inward and fall back to Mars over 5 million years.

Nature Geosci. **http://dx.doi.org/10.1038/ngeo2742 (2016)**

CLIMATE SCIENCE

Warming shifts plant sex ratio

Climate change seems to be skewing the sex ratios of an alpine herb towards male plants.

William Petry at the University of California, Irvine, and his colleagues analysed data on populations of the herb valerian (*Valeriana edulis*) in the Rocky Mountains of Colorado as the region became warmer and drier over the past few decades. They found that in 2011, plants at the highest elevations were only 23% male, whereas at lower altitudes, where the climate is warmer and wetter, the plants were up to 50% male. Across 9 populations at a variety of elevations, there was an average of 6% more males in 2011 than in 1978.

A higher male-to-female ratio could result in increased pollination — and therefore seed production — which could help the plants to expand their range as they adapt to climate change, the authors suggest.

Science 353, **69–71 (2016)**

ROBOTICS

Soft wheels make robots tough

Wheels built entirely from soft materials can help robots to roll over tricky terrain and resist damage.

Aaron Mazzeo and his co-workers at Rutgers University in Piscataway, New Jersey, built a squishy wheel inspired by the inching motions of soft creatures such as earthworms. A stretchable ring contains multiple internal chambers, groups of which can be inflated and deflated sequentially around the circle. The pressurized compartments exert torque on a second, outer ring, causing it to turn.

A soft robotic vehicle fitted with four of these wheels (**pictured**) travelled on a flat surface at 3.7 centimetres per second and kept moving after being dropped from eight times its height. The researchers also drove the robot over a rocky landscape and underwater, and show that their concept can be modified to make winch rotors.

Adv. Mater. **http://doi.org/f3qjsh (2016)**



WILEY

SOCIAL SELECTION

Fake article webpages draw fire

A debate is swirling around a tactic that academic publisher John Wiley & Sons uses to fight online piracy (see go.nature.com/299xi1y). The company created a webpage, accessible by several URLs, that looked like an academic paper to automated downloading programs. But any users who accessed the URLs were then blocked from viewing other Wiley content. Wiley and other publishers use these 'trap' URLs — which are invisible to most human users — to detect and prevent unauthorized downloading and republishing of their content. But some users say that the tactic is too heavy-handed.

NATURE.COM
For more on popular papers:
go.nature.com/29hbqog

CANCER BIOLOGY

Leukaemia cells hide in fat tissue

Cancer-causing stem cells evade chemotherapy by surviving in fat deposits around gonads.

Fat tissue supports the growth of normal blood-forming stem cells. Craig Jordan of the University of Colorado Denver and his colleagues found that in a mouse model of one form of leukaemia, gonadal fat deposits contained numerous cancer stem cells, but subcutaneous fat had very few. Leukaemic cells induced the breakdown of gonadal fat, releasing nutrients that fuelled the growth of malignant cells in fat as well as other tissues. The cancer stem cells also expressed CD36, a cellular marker that boosts fat metabolism, helping to protect the cells from many chemotherapy drugs.

Targeting fat metabolism could help to eradicate leukaemia stem cells, the authors suggest.

Cell Stem Cell **http://doi.org/bkqj (2016)**

CLIMATE CHANGE

Negative carbon emissions needed

Countries' existing promises regarding emissions reductions are unlikely to prevent global warming exceeding 2 °C above pre-industrial temperatures by the end of the century, meaning that large amounts of carbon may need to be removed from the atmosphere.

Benjamin Sanderson and his co-workers at the US National Center for Atmospheric Research in Boulder, Colorado, explored the odds of staying below 2 °C of warming for a range of emissions pathways. They also analysed whether 'negative emissions' — the removal of carbon from the atmosphere — will be necessary.

To avoid crossing the 2-degree threshold during this century, net global emissions must drop to zero by 2085, the authors find. Depending on the level of near-term reductions, between 1.5 billion and 5 billion tonnes of CO_2 per year will need to be captured and removed from the atmosphere thereafter.

Geophys. Res. Lett. **http://doi.org/bkqh (2016)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

© 2016 Macmillan Publishers Limited. All rights reserved.

LETTER

doi:10.1038/nature18627

The quiescent intracluster medium in the core of the Perseus cluster

The Hitomi collaboration*

Clusters of galaxies are the most massive gravitationally bound objects in the Universe and are still forming. They are thus important probes[1] of cosmological parameters and many astrophysical processes. However, knowledge of the dynamics of the pervasive hot gas, the mass of which is much larger than the combined mass of all the stars in the cluster, is lacking. Such knowledge would enable insights into the injection of mechanical energy by the central supermassive black hole and the use of hydrostatic equilibrium for determining cluster masses. X-rays from the core of the Perseus cluster are emitted by the 50-million-kelvin diffuse hot plasma filling its gravitational potential well. The active galactic nucleus of the central galaxy NGC 1275 is pumping jetted energy into the surrounding intracluster medium, creating buoyant bubbles filled with relativistic plasma. These bubbles probably induce motions in the intracluster medium and heat the inner gas, preventing runaway radiative cooling—a process known as active galactic nucleus feedback[2–6]. Here we report X-ray observations of the core of the Perseus cluster, which reveal a remarkably quiescent atmosphere in which the gas has a line-of-sight velocity dispersion of 164 ± 10 kilometres per second in the region 30–60 kiloparsecs from the central nucleus. A gradient in the line-of-sight velocity of 150 ± 70 kilometres per second is found across the 60-kiloparsec image of the cluster core. Turbulent pressure support in the gas is four per cent of the thermodynamic pressure, with large-scale shear at most doubling this estimate. We infer that a total cluster mass determined from hydrostatic equilibrium in a central region would require little correction for turbulent pressure.

The JAXA Hitomi X-ray Observatory[7] was launched on 2016 February 17 from Tanegashima, Japan. It carries the non-dispersive soft X-ray spectrometer (SXS)[8], which is a calorimeter cooled to 0.05 K giving a Gaussian-shaped energy response with a 4.9-eV full-width at half-maximum (FWHM; ratio of photon energy to FWHM, $E/\mathrm{d}E = 1{,}250$ at 6 keV) over a 6 × 6 pixel array (total 3 arcmin × 3 arcmin). It operates over an energy range of 0.3–12 keV with X-rays focused by a mirror[9] with angular resolution of 1.2 arcmin (half-power diameter). A gate valve was in place for early observations to minimize the risk of contamination from outgassing of the spacecraft. It includes a beryllium window that absorbs most X-rays below about 3 keV. The SXS can detect bulk and turbulent motions of the intracluster medium by measuring Doppler shifts and broadening of the emission lines with unprecedented accuracy. It also allows the detection of weak emission lines or absorption features.

The SXS imaged a 60 kpc × 60 kpc region in the Perseus cluster centred 1 arcmin to the northwest of the nucleus for a total exposure time of 230 ks. The offset from the nucleus was due to the attitude control system not having then been calibrated. For this early observation, not all calibration procedures were available; in particular, we did not have contemporaneous calibration of the energy scale factors (gains) of the detector pixels. Gain variation over short time intervals was corrected using a separate calibration pixel illuminated by 5.9-keV Mn Kα photons from an ^{55}Fe X-ray source. Gain values were pinned to an absolute scale via extrapolation of a subsequent calibration of the whole array 10 days later using illumination by another ^{55}Fe source mounted on the filter wheel. (For more details, see Methods.) We used a subset of the Perseus data closest to that calibration to derive the velocity map. For the line-width determination, we used the full dataset to minimize the statistical uncertainty, and applied a scale factor to force the Fe XXV Heα complex from the cluster to have the same energy in all pixels. This minimizes the gain uncertainty in the determination of the velocity dispersion, but also removes any true variations of the line-of-sight velocity of the intracluster medium across the field.

A 5.5–8.5-keV spectrum of the full 3 arcmin × 3 arcmin field is shown in Fig. 1. This spectrum shows a thermal continuum with line emission[10] from Cr, Mn, Fe and Ni. The strongest lines are from Fe and consist of Fe XXV Heα, Heβ and Heγ complexes, together with H-like Fe XXVI Lyman α (Lyα) lines. The total number of counts in the Fe XXV Heα line is 21,726, of which about 16 counts are expected from residual instrumental background. The line complex is spread over about 75 eV and its major components include the resonance, intercombination and forbidden lines, all of which have been resolved.

We adopt a minimally model-dependent method for spectral fitting, and represent the Fe XXV Heα, Fe XXV Heβ and Fe XXVI Lyα complexes in the spectrum with a set of Gaussians with free normalizations and energies fixed either at redshifted laboratory values (for He-like Fe) or theoretical values (for H-like Fe); see Extended Data Table 1. Figure 2 shows the profiles of these lines in a spectrum obtained from the outer region of the Perseus core, which excludes the active galactic nucleus (AGN) and prominent inner bubbles (Fig. 3). To measure the line-of-sight velocity broadening (Gaussian σ), we fitted the high-signal-to-noise, Fe XXV Heα line complex using nine Gaussians associated with lines known from atomic physics and obtain $164 \pm 10\,\mathrm{km\,s^{-1}}$ (all uncertainties are quoted at the 90% confidence level). The widths of the 6.7008-keV resonance line and the 6.617-keV blend of faint satellite lines are allowed to be separate from the rest of the lines. The effect of the thermal broadening expected from the observed 4-keV plasma has been removed (alone it corresponds to $80\,\mathrm{km\,s^{-1}}$). Conservative estimates of the uncertainty in energy resolution (FWHM of 5 ± 0.5 eV) result in a systematic uncertainty range in the turbulent velocity of only $\pm 6\,\mathrm{km\,s^{-1}}$, because the total measured line width is roughly twice the instrumental broadening, which adds to the astronomical broadening in quadrature. Uncertainties in plasma temperature add only a further $\pm 2\,\mathrm{km\,s^{-1}}$. The statistical scatter in the energy-scale alignment of co-added pixels results in an overestimate of the true broadening by not more than $3\,\mathrm{km\,s^{-1}}$. The finite angular resolution of the telescope in the presence of a velocity gradient across the cluster results in a small artificial increase in the measured dispersion (see Methods) that is difficult to quantify at this stage.

The Fe XXVI Lyα complex alone (554 counts) yields a consistent velocity broadening of $160 \pm 16\,\mathrm{km\,s^{-1}}$. A search for spatial variations in velocity broadening using the Fe XXV Heα lines reveals that

*A list of participants and their affiliations appears at the end of the paper.

© 2016 Macmillan Publishers Limited. All rights reserved

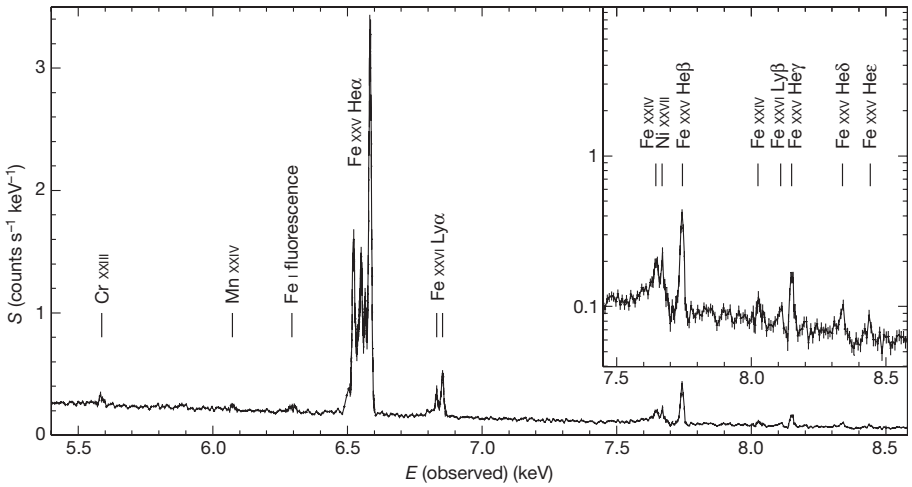


Figure 1 | Full array spectrum of the core of the Perseus cluster obtained by the Hitomi observatory. The redshift of the Perseus cluster is $z = 0.01756$. The inset has a logarithmic scale, which allows the weaker lines to be better seen. The flux S is plotted against photon energy E.

all 1-arcmin-resolution bins have broadening of less than 200 km s^{-1}. With just a single observation we cannot comment on how this result translates to the wider cluster core.

The tightest previous constraint on the velocity dispersion of a cluster gas was from the XMM-Newton reflection grating spectrometer, giving[11,12] an upper limit of 235 km s^{-1} on the X-ray coolest gas (that is, $kT < 3$ keV, where k is Boltzmann's constant and T is the temperature) in the distant luminous cluster A1835. These measurements are available for only a few peaked clusters[13]; the angular size of Perseus and many other bright clusters is too large to derive meaningful velocity results from a slitless dispersive spectrometer such as the reflection grating spectrometer (the corresponding limit for Perseus[13] is 625 km s^{-1}). The Hitomi SXS achieves much higher accuracy on diffuse hot gas owing to it being non-dispersive.

We measure a slightly higher velocity broadening, 187 ± 13 km s^{-1}, in the central region (Fig. 3a) that includes the bubbles and the nucleus. This region exhibits a strong power-law component from the AGN, which is several times brighter than the measurement[14] made in 2001 with XMM-Newton, consistent with the luminosity increase seen at other wavelengths. A fluorescent line from neutral Fe is present in the spectrum (Fig. 1), which can be emitted by the AGN or by the cold gas present in the cluster core[15]. The intracluster medium has a slightly lower average temperature (3.8 ± 0.1 keV) than the outer region (4.1 ± 0.1 keV). By fitting the lines with Gaussians, we measured a ratio of fluxes in Fe XXV Heα resonant and forbidden lines of 2.48 ± 0.16, which is lower than the expected value in optically thin plasma (for $kT = 3.8$ keV, the current APEC[16] and SPEX[17] plasma models give ratios of 2.8 and 2.9–3.6) and suggests the presence of resonant scattering of photons[18]. On the basis of radiative transfer simulations[19] of resonant scattering in these lines, such resonance-line suppression is in broad agreement with that expected for the measured low line widths, providing independent indication of the low level of turbulence. Uncertainties in the current atomic data, as well as more complex structure along the line of sight and across the region, complicate the interpretation of these results, which we defer to a future study.

A velocity map (Fig. 3b) was produced from the absolute energies of the lines in the Fe XXV Heα complex, using a subset of the data for which such a measurement was reliable, given the limited calibration (see Methods). We find a gradient in the line-of-sight velocities of about 150 ± 70 km s^{-1}, from southeast to northwest of the SXS field of view. The velocity to the southeast (towards the nucleus) is 48 ± 17 (statistical) ± 50 (systematic) km s^{-1} redshifted relative to NGC 1275 (redshift $z = 0.01756$) and consistent with results from Suzaku CCD (charge-coupled device) data[20]. Our statistical uncertainty on relative velocities is about 30 times better than that of Suzaku, although there is a systematic uncertainty on the absolute SXS velocities of about 50 km s^{-1} (see Methods).

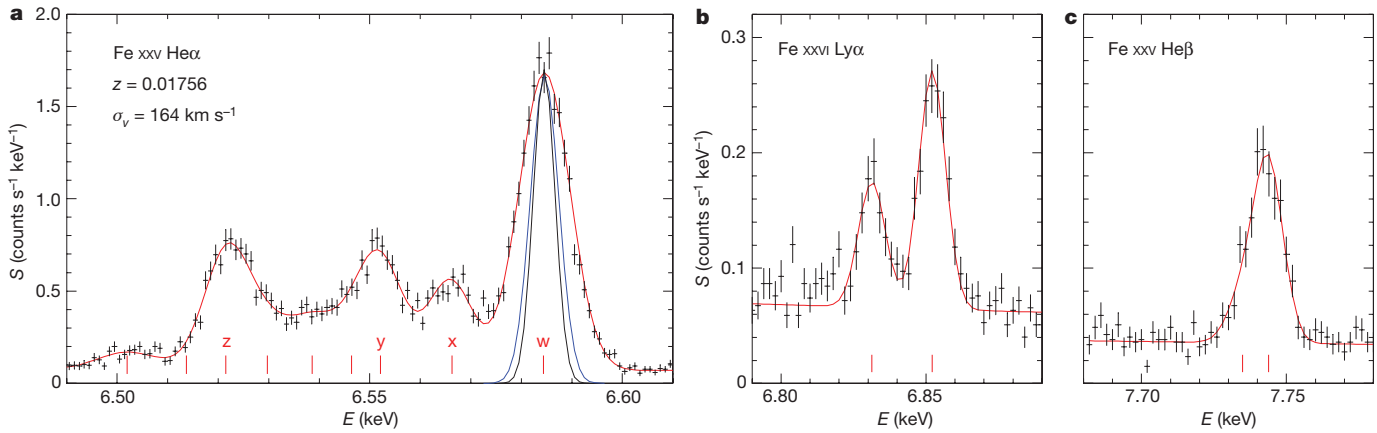


Figure 2 | Spectra of Fe xxv Heα, Fe xxvi Lyα and Fe xxv Heβ from the outer region. a–c, Gaussians (red curves) were fitted to lines with energies (marked by short red lines) from laboratory measurements in the case of He-like Fe xxv (**a**, **c**) and from theory in the case of Fe xxvi Lyα (**b**; see Extended Data Table 1 for details) with the same velocity dispersion ($\sigma_v = 164$ km s^{-1}), except for the Fe xxv Heα resonant line, which was allowed to have its own width. Instrumental broadening with (blue line) and without (black line) thermal broadening are indicated in **a**. The redshift ($z = 0.01756$) is the cluster value to which the data were self-calibrated using the Fe xxv Heα lines. The strongest resonance ('w'), intercombination ('x', 'y') and forbidden ('z') lines are indicated. The error bars are 1 s.d.

© 2016 Macmillan Publishers Limited. All rights reserved

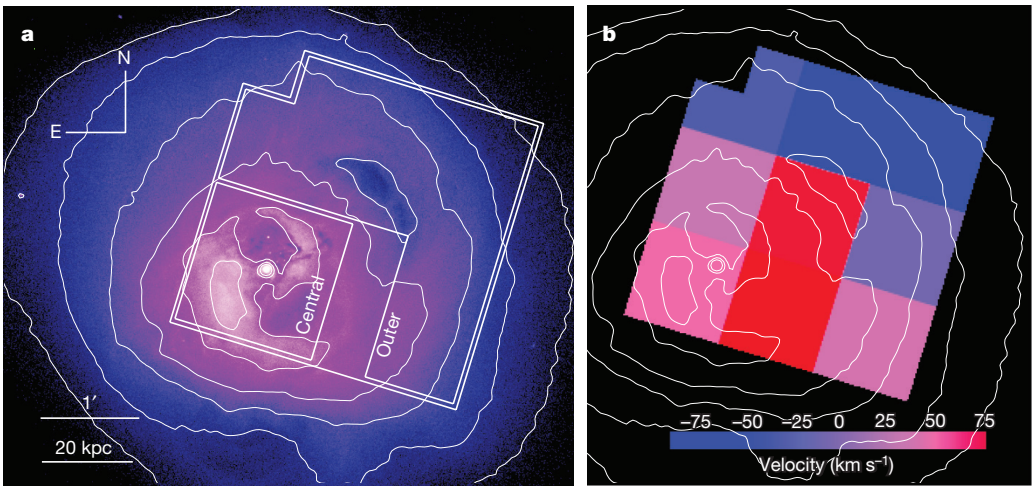


Figure 3 | The region of the Perseus cluster observed by the SXS. **a**, The field of view of the SXS overlaid on a Chandra image. The nucleus of NGC 1275 is seen as the white dot with inner bubbles to the north and south. A buoyant outer bubble lies northwest of the centre of the field. A swirling cold front coincides with the second-most-outer contour. The central and outer regions are marked. **b**, The bulk velocity field across the imaged region. Colours show the difference from the velocity of the central galaxy NGC 1275 (whose redshift is $z = 0.01756$); positive difference means gas receding faster than the galaxy. The 1-arcmin pixels of the map correspond approximately to the angular resolution, but are not entirely independent (see Extended Data Fig. 5). The calibration uncertainty on velocities in individual pixels and in the overall baseline is $50\,\mathrm{km\,s^{-1}}$ ($\Delta z = 0.00017$).

NGC 1275 hosts a giant (80-kpc wide) molecular nebula seen in CO and Hα data with a total cold-gas mass of several $10^{10} M_{\odot}$, which dominates the total gas mass out to a 15-kpc radius. The velocities of that gas[21,22] are consistent with the trend of the SXS bulk shear, suggesting that the molecular gas moves together with the hot plasma. (More details of the X-ray spectra and imaged region are provided in Extended Data Figs 1–8.)

The large-scale bulk shear over the observed 60-kpc field is of comparable amplitude to the small-scale velocity dispersion that we derive for the outer region. The dispersion can be due to gas flows around the rising bubble at the centre of the field[23,24], a velocity gradient in the cold front[25] contained in this region, sound waves[26,27], turbulence[28] or galaxy motions[29]. The large-scale shear could be due to the buoyant AGN bubbles or to sloshing motions of gas in the cluster core that give rise to the cold front[25].

If the observed dispersion is interpreted as turbulence driven on scales comparable with the size of the largest bubbles in the field (about 20–30 kpc), then it is in agreement with the level inferred[28] from X-ray surface brightness fluctuations. In this case, our measured velocity dispersion suggests that turbulent dissipation of kinetic energy can offset radiative cooling. However, assuming isotropic turbulence, the ratio of turbulent pressure to thermal pressure in the intracluster medium is low at 4%. Such low-velocity turbulence cannot spread far (<10 kpc) across the cooling core during the fraction (4%) of the cooling time in which it must be replenished, so the turbulent-dissipation mechanism requires that turbulence be generated *in situ* throughout the core. Another process is needed to transport energy from the bubbling region. The observed level of turbulence is also sufficient to sustain the population of ultrarelativistic electrons that give rise to the radio synchrotron mini-halo observed in the Perseus core[30].

A low level of turbulent pressure measured for the core region of a cluster, which is continuously stirred by a central AGN and gas sloshing, is surprising and may imply that turbulence in the intracluster medium is difficult to generate and/or easy to damp. If true throughout the cluster, then this is encouraging for total mass measurements, which depend on knowledge of all forms of pressure support, and for cluster cosmology, which depends on accurate masses.

The Hitomi spacecraft lost its ground contact on 2016 March 26, and later the recovery operation by JAXA was discontinued.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 April; accepted 4 June 2016.

1. Allen, S. W., Evrard, A. E. & Mantz, A. B. Cosmological parameters from observations of galaxy clusters. *Annu. Rev. Astron. Astrophys.* **49,** 409–470 (2011).
2. Boehringer, H., Voges, W., Fabian, A. C., Edge, A. C. & Neumann, D. M. A *ROSAT* HRI study of the interaction of the X-ray-emitting gas and radio lobes of NGC 1275. *Mon. Not. R. Astron. Soc.* **264,** L25–L28 (1993).
3. Churazov, E., Forman, W., Jones, C. & Böhringer, H. Asymmetric, arc minute scale structures around NGC 1275. *Astron. Astrophys.* **356,** 788–794 (2000).
4. McNamara, B. R. *et al. Chandra* X-Ray observations of the Hydra A cluster: an interaction between the radio source and the X-Ray-emitting gas. *Astrophys. J.* **534,** L135–L138 (2000).
5. Fabian, A. C. *et al. Chandra* imaging of the complex X-ray core of the Perseus cluster. *Mon. Not. R. Astron. Soc.* **318,** L65–L68 (2000).
6. Fabian, A. C. Observational evidence of active galactic nuclei feedback. *Annu. Rev. Astron. Astrophys.* **50,** 455–489 (2012).
7. Takahashi, T. *et al.* The ASTRO-H X-ray astronomy satellite. *Proc. SPIE* **9144,** 914425 (2014).
8. Mitsuda, K. *et al.* Soft x-ray spectrometer (SXS): the high-resolution cryogenic spectrometer onboard ASTRO-H. *Proc. SPIE* **9144,** 91442A (2014).
9. Soong, Y. *et al.* ASTRO-H soft X-ray telescope (SXT). *Proc. SPIE* **9144,** 914428 (2014).
10. Tamura, T. *et al.* X-ray spectroscopy of the core of the Perseus cluster with *Suzaku*: elemental abundances in the intracluster medium. *Astrophys. J.* **705,** L62–L66 (2009).
11. Sanders, J. S., Fabian, A. C., Smith, R. K. & Peterson, J. R. A direct limit on the turbulent velocity of the intracluster medium in the core of Abell 1835 from *XMM-Newton*. *Mon. Not. R. Astron. Soc.* **402,** L11–L15 (2010).
12. Sanders, J. S., Fabian, A. C. & Smith, R. K. Constraints on turbulent velocity broadening for a sample of clusters, groups and elliptical galaxies using *XMM-Newton*. *Mon. Not. R. Astron. Soc.* **410,** 1797–1812 (2011).
13. Pinto, C. *et al.* Chemical Enrichment RGS cluster Sample (CHEERS): constraints on turbulence. *Astron. Astrophys.* **575,** A38 (2015).
14. Churazov, E., Forman, W., Jones, C. & Böhringer, H. *XMM-Newton* observations of the Perseus cluster. I. The temperature and surface brightness structure. *Astrophys. J.* **590,** 225–237 (2003).
15. Churazov, E., Sunyaev, R., Gilfanov, M., Forman, W. & Jones, C. The 6.4-keV fluorescent iron line from cluster cooling flows. *Mon. Not. R. Astron. Soc.* **297,** 1274–1278 (1998).
16. Foster, A. R., Li, J., Smith, R. K. & Brickhouse, N. S. Updated atomic data and calculations for X-Ray spectroscopy. *Astrophys. J.* **756,** 128 (2012).
17. Kaastra, J. S., Mewe, R. & Nieuwenhuijzen, H. SPEX: a new code for spectral analysis of X & UV spectra. In *11th Colloquium on UV and X-ray Spectroscopy of Astrophysical and Laboratory Plasmas* (eds Yamashita, K. & Watanabe, T.) 411–414 (1996).
18. Gil'fanov, M. R., Syunyaev, R. A. & Churazov, E. M. Radial brightness profiles of resonance x-ray lines in galaxy clusters. *Sov. Astron. Lett.* **13,** 3–7 (1987).

© 2016 Macmillan Publishers Limited. All rights reserved

19. Zhuravleva, I. *et al.* Resonant scattering in the Perseus cluster: spectral model for constraining gas motions with *Astro-H*. *Mon. Not. R. Astron. Soc.* **435,** 3111–3121 (2013).
20. Tamura, T. *et al.* Gas bulk motion in the Perseus cluster measured with *Suzaku*. *Astrophys. J.* **782,** 38 (2014).
21. Salomé, P. *et al.* A very extended molecular web around NGC 1275. *Astron. Astrophys.* **531,** A85 (2011).
22. Hatch, N. A., Crawford, C. S., Johnstone, R. M. & Fabian, A. C. On the origin and excitation of the extended nebula surrounding NGC 1275. *Mon. Not. R. Astron. Soc.* **367,** 433–448 (2006).
23. Bruggen, M., Hoeft, M. & Ruszkowski, M. X-Ray line tomography of AGN-induced motion in clusters of galaxies. *Astrophys. J.* **628,** 153–159 (2005).
24. Heinz, S., Bruggen, M. & Morsony, B. Prospects of high-resolution X-ray spectroscopy for active galactic nucleus feedback in galaxy clusters. *Astrophys. J.* **708,** 462–468 (2010).
25. Markevitch, M. & Vikhlinin, A. Shocks and cold fronts in galaxy clusters. *Phys. Rep.* **443,** 1–53 (2007).
26. Fabian, A. C. *et al.* A deep *Chandra* observation of the Perseus cluster: shocks and ripples. *Mon. Not. R. Astron. Soc.* **344,** L43–L47 (2003).
27. Ruszkowski, M., Brüggen, M. & Begelman, M. C. Cluster heating by viscous dissipation of sound waves. *Astrophys. J.* **611,** 158–163 (2004).
28. Zhuravleva, I. *et al.* Turbulent heating in galaxy clusters brightest in X-rays. *Nature* **515,** 85–87 (2014).
29. Gu, L. *et al.* Probing of the interactions between the hot plasmas and galaxies in clusters from $z = 0.1$ to 0.9. *Astrophys. J.* **767,** 157 (2013).
30. ZuHone, J. A., Markevitch, M., Brunetti, G. & Giacintucci, S. Turbulence and radio mini-halos in the sloshing cores of galaxy clusters. *Astrophys. J.* **762,** 78 (2013).

Acknowledgements We acknowledge all the JAXA members who have contributed to the ASTRO-H (Hitomi) project. All US members gratefully acknowledge support through the NASA Science Mission Directorate. Stanford and SLAC members acknowledge support via DoE contract to SLAC National Accelerator Laboratory DE-AC3-76SF00515 and NASA grant NNX15AM19G. Part of this work was performed under the auspices of the US DoE by LLNL under contract DE-AC52-07NA27344 and also supported by NASA grants to LLNL. Support from the European Space Agency is gratefully acknowledged. French members acknowledge support from CNES, the Centre National d'Etudes Spatiales. SRON is supported by NWO, the Netherlands Organization for Scientific Research. The Swiss team acknowledges support of the Swiss Secretariat for Education, Research and Innovation SERI and ESA's PRODEX programme. The Canadian Space Agency is acknowledged for the support of Canadian members. We acknowledge support from JSPS/MEXT KAKENHI grant numbers 15H02070, 15K05107, 23340071, 26109506, 24103002, 25400236, 25800119, 25400237, 25287042, 24540229, 25105516, 23540280, 25400235, 25247028, 26800095, 25400231, 25247028, 26220703, 24105007, 23340055, 15H00773, 23000004, 15H02090, 15K17610, 15H05438, 15H00785 and 24540232. H. Akamatsu acknowledges support of NWO via a Veni grant. M. Axelsson acknowledges a JSPS International Research Fellowship. C. Done acknowledges STFC funding under grant ST/L00075X/1. P. Gandhi acknowledges a JAXA International Top Young Fellowship and UK Science and Technology Funding Council (STFC) grant ST/J003697/2. H. Russell, A. C. Fabian and C. Pinto acknowledge support from ERC Advanced Grant Feedback 340442. We thank contributions by many companies, including, in particular, NEC, Mitsubishi Heavy Industries, Sumitomo Heavy Industries and Japan Aviation Electronics Industry. Finally, we acknowledge strong support from the following engineers. JAXA/ISAS: C. Baluta, N. Bando, A. Harayama, K. Hirose, K. Ishimura, N. Iwata, T. Kawano, S. Kawasaki, K. Minesugi, C. Natsukari, H. Ogawa, M. Ogawa, M. Ohta, T. Okazaki, S.-i. Sakai, Y. Shibano, M. Shida, T. Shimada, A. Wada, T. Yamada; JAXA/TKSC: A. Okamoto, Y. Sato, K. Shinozaki, H. Sugita; Chubu U: Y. Namba; Ehime U: K. Ogi; Kochi U of Technology: T. Kosaka; Miyazaki U: Y. Nishioka; Nagoya U: H. Nagano; NASA/GSFC: T. Bialas, K. Boyce, E. Canavan, M. DiPirro, M. Kimball, C. Masters, D. Mcguinness, J. Miko, T. Muench, J. Pontius, P. Shirron, C. Simmons, G. Sneiderman, T. Watanabe; Noqsi Aerospace Ltd: J. Doty; Stanford U/KIPAC: M. Asai, K. Gilmore; ESA (Netherlands): C. Jewell; SRON: D. Haas, M. Frericks, P. Laubert, P. Lowes; U of Geneva: P. Azzarello; CSA: A. Koujelev, F. Moroso.

Author Contributions The science goals of Hitomi (known as ASTRO-H before launch) were discussed and developed over more than 10 years by the ASTRO-H Science Working Group (SWG), all members of which are authors of this manuscript. All the instruments were prepared by joint efforts of the team. Calibration of the Perseus dataset was carried out by members of the SXS team. Data analysis and manuscript preparation were carried out by a small subgroup of authors appointed by the SWG, on the basis of the extensive discussion made in the white paper produced by all SWG members. The manuscript was subject to an internal collaboration-wide review process. All authors reviewed and approved the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A. C. Fabian (acf@ast.cam.ac.uk).

Hitomi Collaboration

Felix Aharonian[1,2], Hiroki Akamatsu[3], Fumie Akimoto[4], Steven W. Allen[5,6,7], Naohisa Anabuki[8], Lorella Angelini[9], Keith Arnaud[9,10], Marc Audard[11], Hisamitsu Awaki[12], Magnus Axelsson[13], Aya Bamba[14], Marshall Bautz[15], Roger Blandford[5,6,7], Laura Brenneman[16], Gregory V. Brown[17], Esra Bulbul[15], Edward Cackett[18], Maria Chernyakova[1], Meng Chiao[9], Paolo Coppi[19], Elisa Costantini[3], Jelle de Plaa[3], Jan-Willem den Herder[3], Chris Done[20], Tadayasu Dotani[21], Ken Ebisawa[21], Megan Eckart[9], Teruaki Enoto[22,23], Yuichiro Ezoe[13], Andrew C. Fabian[18], Carlo Ferrigno[11], Adam Foster[16], Ryuichi Fujimoto[24], Yasushi Fukazawa[25], Akihiro Furuzawa[4], Massimiliano Galeazzi[26], Luigi Gallo[27], Poshak Gandhi[28], Margherita Giustini[3], Andrea Goldwurm[29], Liyi Gu[3], Matteo Guainazzi[21,30], Yoshito Haba[31], Kouichi Hagino[21], Kenji Hamaguchi[9,32], Ilana Harrus[9,32], Isamu Hatsukade[33], Katsuhiro Hayashi[21], Takayuki Hayashi[4], Kiyoshi Hayashida[8], Junko Hiraga[34], Ann Hornschemeier[9], Akio Hoshino[35], John Hughes[36], Ryo Iizuka[21], Hajime Inoue[21], Yoshiyuki Inoue[21], Kazunori Ishibashi[4], Manabu Ishida[21], Kumi Ishikawa[37], Yoshitaka Ishisaki[13], Masayuki Itoh[38], Naoko Iyomoto[39], Jelle Kaastra[3], Timothy Kallman[9], Tuneyoshi Kamae[5], Erin Kara[10], Jun Kataoka[40], Satoru Katsuda[41], Junichiro Katsuta[25], Madoka Kawaharada[42], Nobuyuki Kawai[43], Richard Kelley[9], Dmitry Khangulyan[35], Caroline Kilbourne[9], Ashley King[5,6], Takao Kitaguchi[25], Shunji Kitamoto[35], Tetsu Kitayama[44], Takayoshi Kohmura[45], Motohide Kokubun[21], Shu Koyama[21], Katsuji Koyama[46], Peter Kretschmar[30], Hans Krimm[9,47], Aya Kubota[48], Hideyo Kunieda[4], Philippe Laurent[29], François Lebrun[29], Shiu-Hang Lee[21], Maurice Leutenegger[9], Olivier Limousin[29], Michael Loewenstein[9,10], Knox S. Long[49], David Lumb[50], Grzegorz Madejski[5,7], Yoshitomo Maeda[21], Daniel Maier[29], Kazuo Makishima[51], Maxim Markevitch[9], Hironori Matsumoto[52], Kyoko Matsushita[53], Dan McCammon[54], Brian McNamara[55], Missagh Mehdipour[3], Eric Miller[15], Jon Miller[56], Shin Mineshige[22], Kazuhisa Mitsuda[21], Ikuyuki Mitsuishi[4], Takuya Miyazawa[4], Tsunefumi Mizuno[25], Hideyuki Mori[9], Koji Mori[33], Harvey Moseley[9], Koji Mukai[9,32], Hiroshi Murakami[57], Toshio Murakami[24], Richard Mushotzky[10], Ryo Nagino[8], Takao Nakagawa[21], Hiroshi Nakajima[8], Takeshi Nakamori[58], Toshio Nakano[37], Shinya Nakashima[21], Kazuhiro Nakazawa[14], Masayoshi Nobukawa[59], Hirofumi Noda[37], Masaharu Nomachi[60], Steve O'Dell[61], Hirokazu Odaka[21], Takaya Ohashi[13], Masanori Ohno[25], Takashi Okajima[9], Naomi Ota[62], Masanobu Ozaki[21], Frits Paerels[63], Stephane Paltani[11], Arvind Parmar[50], Robert Petre[9], Ciro Pinto[18], Martin Pohl[11], F. Scott Porter[9], Katja Pottschmidt[9,32], Brian Ramsey[61], Christopher Reynolds[10], Helen Russell[18], Samar Safi-Harb[64], Shinya Saito[35], Kazuhiro Sakai[9], Hiroaki Sameshima[21], Goro Sato[21], Kosuke Sato[53], Rie Sato[21], Makoto Sawada[65], Norbert Schartel[30], Peter Serlemitsos[9], Hiromi Seta[13], Megumi Shidatsu[51], Aurora Simionescu[21], Randall Smith[16], Yang Soong[9], Lukasz Stawarz[66], Yasuharu Sugawara[41], Satoshi Sugita[43], Andrew Szymkowiak[19], Hiroyasu Tajima[67], Hiromitsu Takahashi[25], Tadayuki Takahashi[21], Shin'ichiro Takeda[68], Yoh Takei[21], Toru Tamagawa[37], Keisuke Tamura[4], Takayuki Tamura[21], Takaaki Tanaka[46], Yasuo Tanaka[21], Yasuyuki Tanaka[25], Makoto Tashiro[69], Yuzuru Tawara[4], Yukikatsu Terada[69], Yuichi Terashima[12], Francesco Tombesi[9], Hiroshi Tomida[21], Yohko Tsuboi[41], Masahiro Tsujimoto[21], Hiroshi Tsunemi[8], Takeshi Tsuru[46], Hiroyuki Uchida[46], Hideki Uchiyama[70], Yasunobu Uchiyama[35], Shutaro Ueda[21], Yoshihiro Ueda[22], Shiro Ueno[21], Shin'ichiro Uno[71], Meg Urry[19], Eugenio Ursino[26], Cor de Vries[3], Shin Watanabe[21], Norbert Werner[5,6], Daniel Wik[9,72], Dan Wilkins[27], Brian Williams[9], Shinya Yamada[13], Hiroya Yamaguchi[9], Kazutaka Yamaoka[4], Noriko Y. Yamasaki[21], Makoto Yamauchi[33], Shigeo Yamauchi[62], Tahir Yaqoob[9,32], Yoichi Yatsu[43], Daisuke Yonetoku[24], Atsumasa Yoshida[65], Takayuki Yuasa[37], Irina Zhuravleva[5,6] & Abderahmen Zoghbi[56]

[1]Astronomy and Astrophysics Section, Dublin Institute for Advanced Studies, Dublin 2, Ireland. [2]National Research Nuclear University (MEPHI), 115409 Moscow, Russia. [3]SRON Netherlands Institute for Space Research, Utrecht, The Netherlands. [4]Department of Physics, Nagoya University, Aichi 464-8602, Japan. [5]Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, California 94305, USA. [6]Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, California 94305, USA. [7]SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. [8]Department of Earth and Space Science, Osaka University, Osaka 560-0043, Japan. [9]NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [10]Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. [11]Université de Genève, 1211 Genève 4, Switzerland. [12]Department of Physics, Ehime University, Ehime 790-8577, Japan. [13]Department of Physics, Tokyo Metropolitan University, Tokyo 192-0397, Japan. [14]Department of Physics, University of Tokyo, Tokyo 113-0033, Japan. [15]Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [16]Smithsonian Astrophysical Observatory, 60 Garden Street, MS-4, Cambridge, Massachusetts 02138, USA. [17]Lawrence Livermore National Laboratory, Livermore, California 94550, USA. [18]Institute of Astronomy, Cambridge University, Cambridge CB3 0HA, UK. [19]Yale Center for Astronomy and Astrophysics, Yale University, New Haven, Connecticut 06520-8121, USA. [20]Department of Physics, University of Durham, Durham DH1 3LE, UK. [21]Institute of Space and Astronautical Science (ISAS), Japan Aerospace Exploration Agency (JAXA), Kanagawa 252-5210, Japan. [22]Department of Astronomy, Kyoto University, Kyoto 606-8502, Japan. [23]The Hakubi Center for Advanced Research, Kyoto University, Kyoto 606-8302, Japan. [24]Faculty of Mathematics and Physics, Kanazawa University, Ishikawa 920-1192, Japan. [25]Department of Physical Science, Hiroshima University, Hiroshima 739-8526, Japan. [26]Physics Department, University of Miami, Miami, Florida 33124, USA. [27]Department of Astronomy and Physics, Saint Mary's University,

© 2016 Macmillan Publishers Limited. All rights reserved

Halifax, Nova Scotia B3H 3C3, Canada. [28]Department of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK. [29]IRFU/Service d'Astrophysique, CEA Saclay, 91191 Gif-sur-Yvette Cedex, France. [30]European Space Agency (ESA), European Space Astronomy Centre (ESAC), Madrid, Spain. [31]Department of Physics and Astronomy, Aichi University of Education, Aichi 448-8543, Japan. [32]Department of Physics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250, USA. [33]Department of Applied Physics and Electronic Engineering, University of Miyazaki, Miyazaki 889-2192, Japan. [34]Department of Physics, School of Science and Technology, Kwansei Gakuin University, Hyogo 669-1337, Japan. [35]Department of Physics, Rikkyo University, Tokyo 171-8501, Japan. [36]Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854-8019, USA. [37]RIKEN Nishina Center, Saitama 351-0198, Japan. [38]Faculty of Human Development, Kobe University, Hyogo 657-8501, Japan. [39]Kyushu University, Fukuoka 819-0395, Japan. [40]Research Institute for Science and Engineering, Waseda University, Tokyo 169-8555, Japan. [41]Department of Physics, Chuo University, Tokyo 112-8551, Japan. [42]Tsukuba Space Center (TKSC), Japan Aerospace Exploration Agency (JAXA), Ibaraki 305-8505, Japan. [43]Department of Physics, Tokyo Institute of Technology, Tokyo 152-8551, Japan. [44]Department of Physics, Toho University, Chiba 274-8510, Japan. [45]Department of Physics, Tokyo University of Science, Chiba 278-8510, Japan. [46]Department of Physics, Kyoto University, Kyoto 606-8502, Japan. [47]Universities Space Research Association, 7178 Columbia Gateway Drive, Columbia, Maryland 21046, USA. [48]Department of Electronic Information Systems, Shibaura Institute of Technology, Saitama 337-8570, Japan. [49]Space Telescope Science Institute, Baltimore, Maryland 21218, USA. [50]European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), 2200 AG Noordwijk, The Netherlands. [51]RIKEN, Saitama 351-0198, Japan. [52]Kobayashi-Maskawa Institute, Nagoya University, Aichi 464-8602, Japan. [53]Department of Physics, Tokyo University of Science, Tokyo 162-8601, Japan. [54]Department of Physics, University of Wisconsin, Madison, Wisconsin 53706, USA. [55]University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. [56]Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA. [57]Department of Information Science, Faculty of Liberal Arts, Tohoku Gakuin University, Miyagi 981-3193, Japan. [58]Department of Physics, Faculty of Science, Yamagata University, Yamagata 990-8560, Japan. [59]Department of Teacher Training and School Education, Nara University of Education, Takabatake-cho, Nara 630-8528, Japan. [60]Research Center for Nuclear Physics (Toyonaka), Osaka University, 1-1 Machikaneyama-machi, Toyonaka, Osaka 560-0043, Japan. [61]NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA. [62]Department of Physics, Faculty of Science, Nara Women's University, Nara 630-8506, Japan. [63]Department of Astronomy, Columbia University, New York, New York 10027, USA. [64]Department of Physics and Astronomy, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. [65]Department of Physics and Mathematics, Aoyama Gakuin University, Kanagawa 252-5258, Japan. [66]Astronomical Observatory, Jagiellonian University, 30-244 Kraków, Poland. [67]Institute of Space-Earth Environmental Research, Nagoya University, Aichi 464-8601, Japan. [68]Advanced Medical Instrumentation Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Okinawa 904-0495, Japan. [69]Department of Physics, Saitama University, Saitama 338-8570, Japan. [70]Science Education, Faculty of Education, Shizuoka University, Shizuoka 422-8529, Japan. [71]Faculty of Health Sciences, Nihon Fukushi University, Aichi 475-0012, Japan. [72]Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Gain corrections and calibration. Gain scales for each pixel were measured in ground calibration using a series of fiducial X-ray lines at several detector heat-sink temperatures (a single spectral energy reference is sufficient to determine the effective detector temperature and thus the appropriate gain curve to use). As the heat-sink temperature varies, the gain of each pixel tracks the gain change in the separate calibration pixel that is continuously illuminated by a dedicated ^{55}Fe source. However, time-varying differential thermal loading of the pixels changes their gains by different factors. Thus, use of the gain history of the calibration pixel alone can be insufficient to correct the gain scale of the main array.

The Perseus observation used for this work was performed in two parts, 7 days apart, during which the gain of the calibration pixel changed by 0.6%. Ten days after the final observation, a fiducial measurement for the full array was obtained with an on-board ^{55}Fe source mounted on a filter wheel. To relate this calibration to the two Perseus observations, a two-stage approach was used. First, a correction factor was applied to all pixels using the gain history of the calibration pixel. Second, the differential pixel–pixel gain error was removed using the science observation itself. To do this, the two Perseus observations were subdivided, and the He-like Fe complex was fitted for each pixel in each subset. The time-dependent relative gain of each pixel (compared to the gain correction of the calibration pixel) was then linearly fitted and extrapolated to the later full-array calibration. The full dataset was then corrected using this time-dependent gain function, and the fitting errors were incorporated into the error analysis. To validate this approach, we compared the first observation, which required a substantial gain correction, to the second, for which the instrument was much closer to thermal equilibrium and thus required much less correction. In the first case, the bulk velocity uncertainties are dominated by the uncertainties in the gain correction, whereas, in the second, the uncertainties are dominated by the fit to the He-like Fe complex. The results for the two datasets agree for both bulk velocity and velocity dispersion, indicating that this is a robust approach. For the absolute velocity maps, we are presenting only the result from the second observation of the two used in this work, which requires the least correction and thus has the smallest uncertainty. Note that the limited gain calibration results in pixel-to-pixel uncertainty of $50\,\mathrm{km\,s^{-1}}$ on the absolute velocities.

To derive the absolute velocities of the cluster, we applied the heliocentric correction, which was $-26.4\,\mathrm{km\,s^{-1}}$ for the observation used for velocity mapping. The orbital motion of the satellite around Earth averages out. Our velocities are compared to the heliocentric velocity of NGC 1275 in Fig. 3 and Extended Data Fig. 6.

An additional validation of our calibration comes from a weak background line in the whole-array spectrum from stray ^{55}Fe X-rays, which, after the above procedure, is observed at the correct energy to ± 1.8 eV (equivalent to $\pm 90\,\mathrm{km\,s^{-1}}$). Although the line is not strong enough to verify the calibration of individual pixels (because there should be about 68 counts in this line, non-uniformly distributed across the array), it is a convincing check of the approach.

To determine velocity dispersion, we applied additional scale factors for each SXS pixel to match the apparent energies of the cluster Fe Heα complex in order to remove any residual gain errors at the relevant energy. This also removes the effect of true bulk shear. Pixels were then combined in physically relevant regions to minimize statistical uncertainties.

We presumed a fixed energy resolution of 5.0-eV FWHM in all the analyses. By comparing the line widths in the first and second parts of the observation to estimate the broadening from residual gain drift, and accounting for the variation in resolution of the calibration pixel in time over the observation and during the later calibration of the array, we estimate that the composite resolution of the array and of the separately analysed central and outer regions is bounded with high confidence between 4.5-eV and 5.5-eV FWHM. This 10% uncertainty in instrumental broadening produces a much smaller fractional uncertainty in velocity broadening because the instrumental broadening is roughly half as large as the astronomical broadening, and adds in quadrature with it.

The error from gain aligning the different pixels in a region is smaller than the uncertainty in instrumental broadening because of the small statistical errors in the determination of the scale factor at the Fe Heα complex (in an outer pixel, equivalent to $30\,\mathrm{km\,s^{-1}}$ at 90% confidence). Adding the spectra of multiple pixels with the same velocity uncertainty will add $30\,\mathrm{km\,s^{-1}}$ of noise in quadrature with the measured broadening, producing an overestimate by no more than $3\,\mathrm{km\,s^{-1}}$.

Our velocity dispersion measurements exclude velocity variations across the field on scales of 20 kpc and greater because of the aforementioned self-calibration procedure, but integrate over all scales along the line of sight (weighted by X-ray emissivity, which essentially limits integration to the cluster core). Any comparison with simulations will have to take these into account.

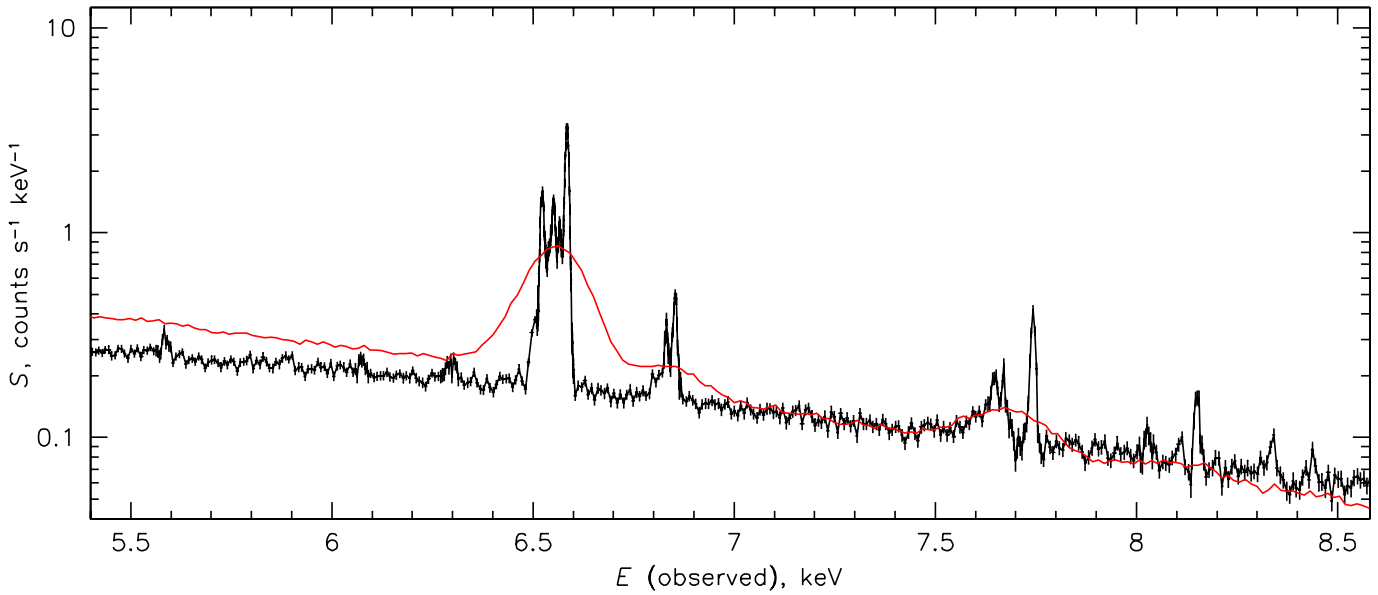
Effects of angular resolution. The point spread function (PSF) of the telescope has a 1.2′ half-power diameter as measured during ground calibration. This means that regions used for spectral extraction get photons not only from the corresponding cluster regions in the sky, but also from the surrounding regions. The PSF image is shown in the right panel of Extended Data Fig. 5, centred on the SXS pixel that contains the cluster peak. By comparing the PSF with the middle panel of Extended Data Fig. 5, which shows the image in the Fe Heα line (which comes mostly from the gas, as opposed to the central AGN), we see that the diffuse emission of the cluster is resolved. However, small regions in the detector, such as the $1' \times 1'$ regions of the velocity map shown in Fig. 3b and Extended Data Fig. 6, are significantly correlated. The fraction of the emission that originates in a given 1′ cluster region and ends up in the corresponding 1′ detector region is 36%–37%, with the rest spreading over the surrounding regions. For example, for the region marked '−60' in Extended Data Fig. 6, the scattered contribution from the neighbouring region marked '78' is 23% of the flux that originates in region −60 itself; the contribution from −60 into 78 is a similar 22% of the flux that originates and stays in 78. Regions adjacent to the brightness peak (which is in region marked '48') are most affected—the region marked '94' has a ratio of photons scattered in from 48 to its own photons of 27%. This means that the true line-of-sight velocity gradients on a 1′ scale have to be steeper than what we measure, but not by much. Scattered flux from an adjacent region with a large velocity difference (for example, from region 78 to region −60) should contribute lines at a different velocity in the spectrum, but such contributions would be very small compared to the observed line-of-sight velocity dispersion of $>160\,\mathrm{km\,s^{-1}}$. Correction of the PSF effects is left for future work.

The PSF scattering also has the subtle effect of inflating our measured value of velocity dispersion. Although the self-calibration procedure that aligns the Fe Heα lines in each pixel (as described above) removes most of the velocity-gradient contribution from the measured velocity dispersion, it does so after the PSF scattering has occurred and mixed the photons from regions with different line-of-sight velocities, so that contribution remains.

Pointing. For this early observation, accurate pointing direction of the spacecraft was not available. We therefore assumed that the observed brightness peak in the SXS image is the AGN in NGC 1275. The resulting uncertainty of the sky coordinates should be less than 15″. The peak of the source determined in short time intervals revealed a small drift of the source in the detector image, within the above coordinate uncertainty. It causes image smearing that is insignificant compared to the effect of PSF scattering.

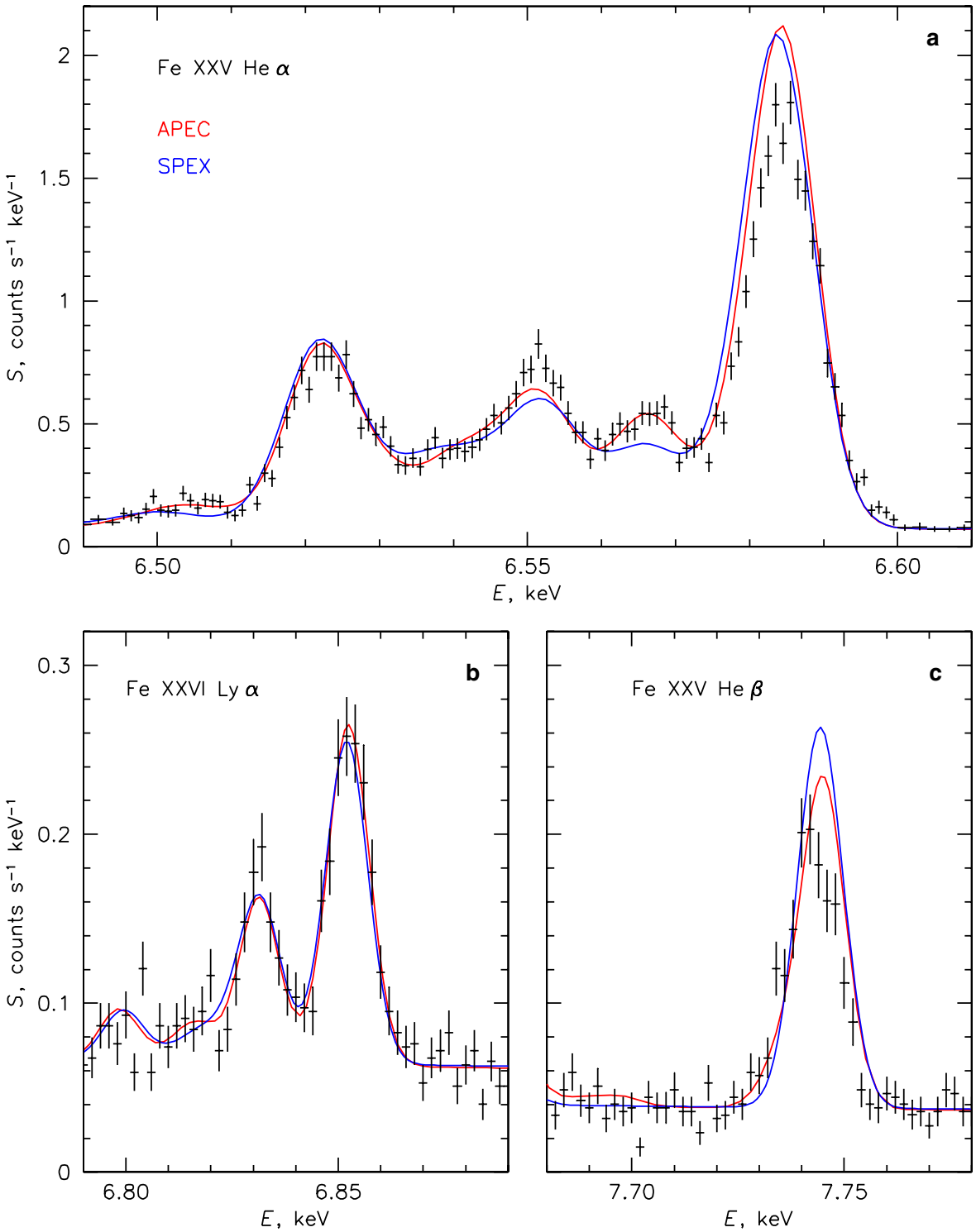
31. Grevesse, N. & Sauval, A. J. Standard solar composition. *Space Sci. Rev.* **85,** 161–174 (1998).
32. Fabian, A. C. *et al.* A wide *Chandra* view of the core of the Perseus cluster. *Mon. Not. R. Astron. Soc.* **418,** 2154–2164 (2011).
33. Ferruit, P., Binette, L., Sutherland, R. S. & Pencontal, E. Tiger observations of the low and high velocity components of NGC 1275. *New Astron.* **2,** 345–363 (1997).
34. Conselice, C., Gallagher, J. G. & Wyse, R. G. On the nature of the NGC 1275 system. *Astron. J.* **122,** 2281–2300 (2001).
35. Beiersdorfer, P. *et al.* High-resolution measurements, line identification, and spectral modeling of Kα transitions in Fe xviii–Fe xxv. *Astrophys. J.* **409,** 846–859 (1993).
36. Smith, A. J. *et al.* Kβ spectra of heliumlike iron from tokomak-fusion-test-reactor plasmas. *Phys. Rev. A* **47,** 3073–3079 (1993).
37. Johnson, W. R. & Soff, G. The lamb shift in hydrogen-like atoms, $1 \leq Z \leq 110$. *Atom. Data Nucl. Data* **33,** 405–446 (1985).

© 2016 Macmillan Publishers Limited. All rights reserved



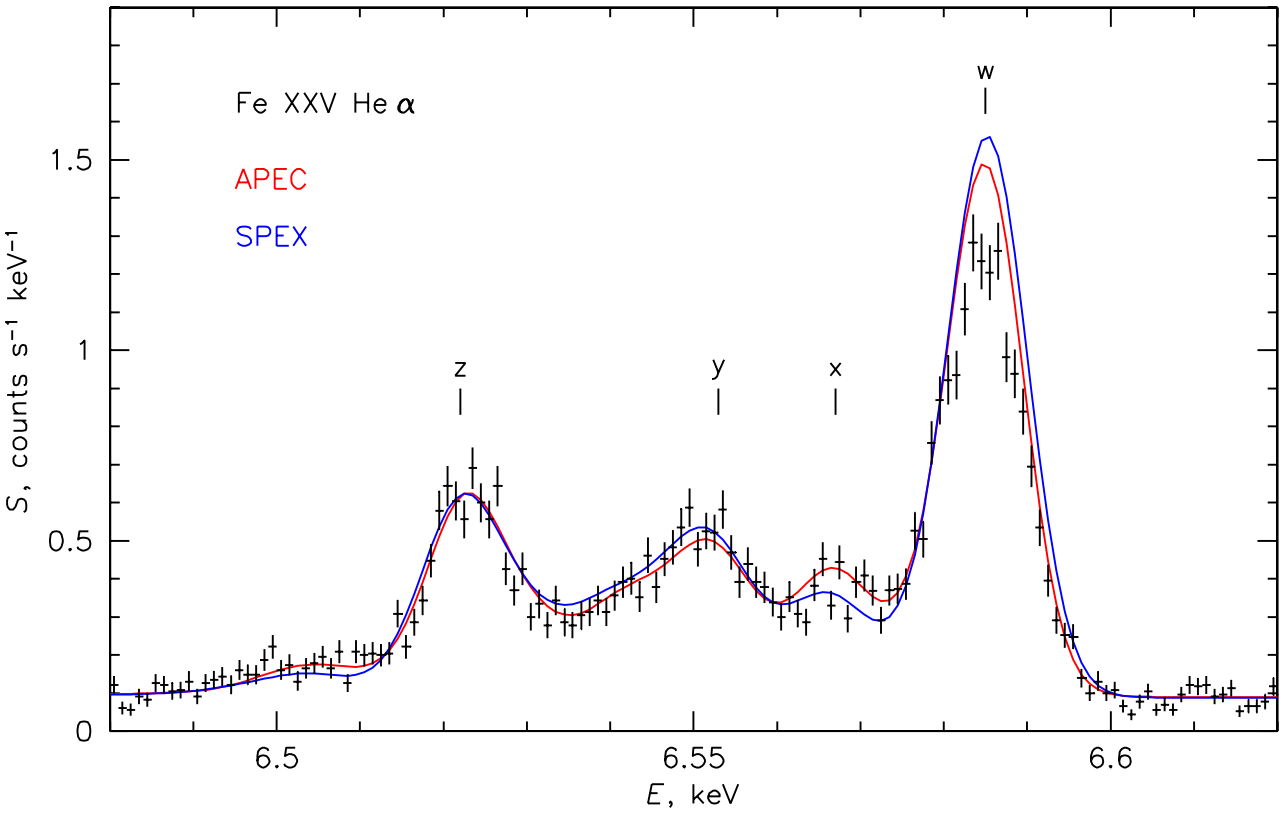
Extended Data Figure 1 | SXS spectrum of the full field overlaid with a CCD spectrum of the same region. The CCD is the Suzaku X-ray imaging spectrometer (XIS) (red line); the difference in the continuum slope is due to differences in the effective areas of the instruments.

© 2016 Macmillan Publishers Limited. All rights reserved



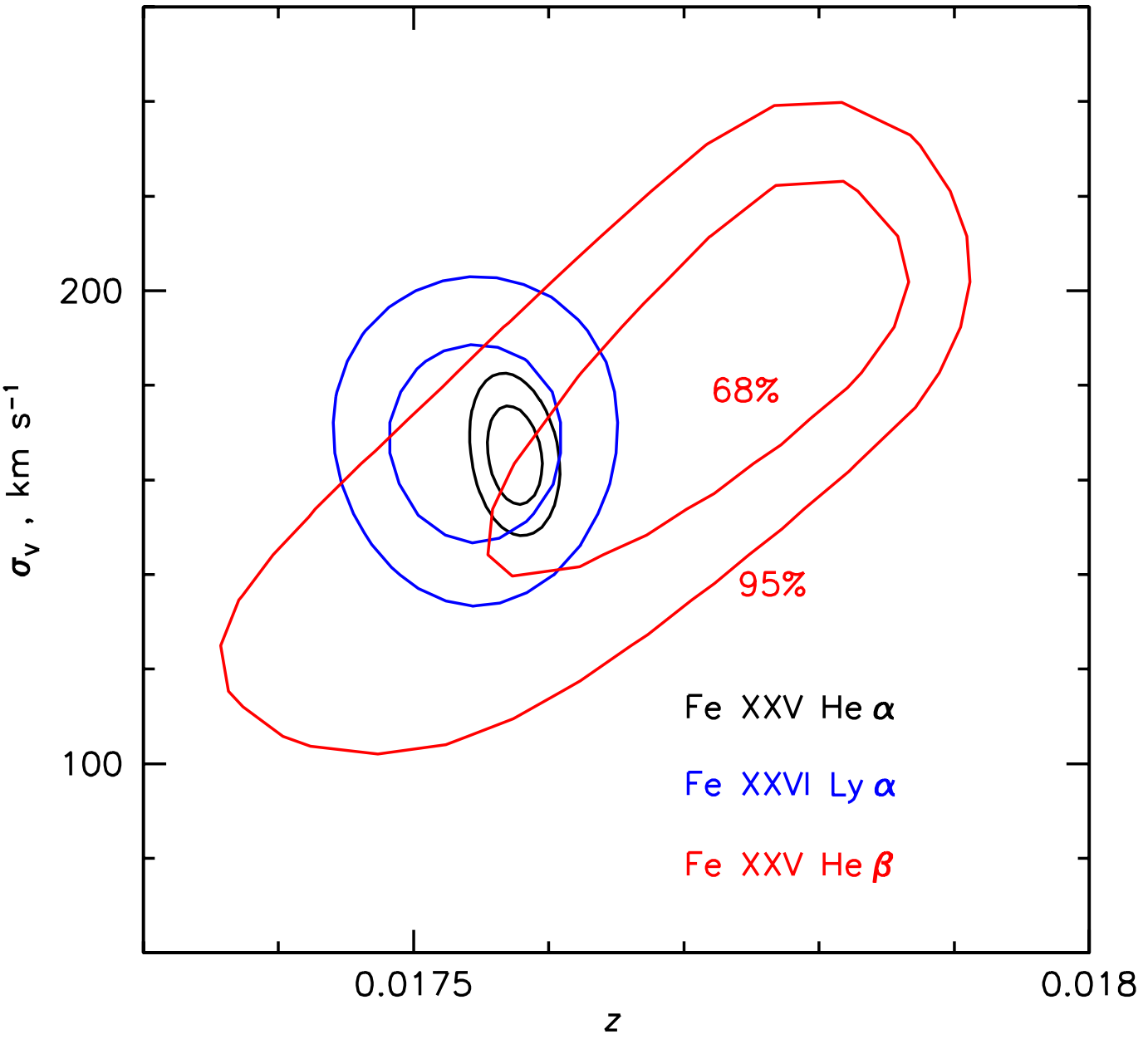
Extended Data Figure 2 | The iron line complexes from the outer region compared with best-fit models. a–c, These have been obtained from various emission-line databases typically used in the literature. The spectra were modelled as a single-temperature, optically thin plasma in collisional ionization equilibrium using either APEC/ATOMDB 3.0.3 (ref 16; red) or SPEX 3.0 (ref. 17; blue). We determined the best-fit model by fitting the Hitomi spectrum from the outer 23 pixels in the energy range 6.4–8 keV, excluding the Fe Heα resonance line and Ni Heα line complex. We obtain consistent best-fit parameters, with both APEC and SPEX predicting a temperature of 4.1 ± 0.1 keV. The iron-to-hydrogen abundances are 0.62 ± 0.02 from APEC and 0.74 ± 0.02 from SPEX, relative to solar values[31]. The line broadening obtained from APEC, 146 ± 7 km s^{-1}, is smaller than the best-fit SPEX value of 171 ± 7 km s^{-1}, although both values are consistent with the line broadening obtained by fitting a set of Gaussians (the result presented in the main body of the paper). Apart from the Fe Heα w line affected by resonance scattering (**a**), both emission line models presented here currently have difficulty reproducing the measured Fe Heα intercombination lines (**a**) as well as the exact position of the Fe Heβ line (**c**). This motivates the model-independent approach adopted in the manuscript for determining the line widths. Error bars are 1 s.d.

© 2016 Macmillan Publishers Limited. All rights reserved



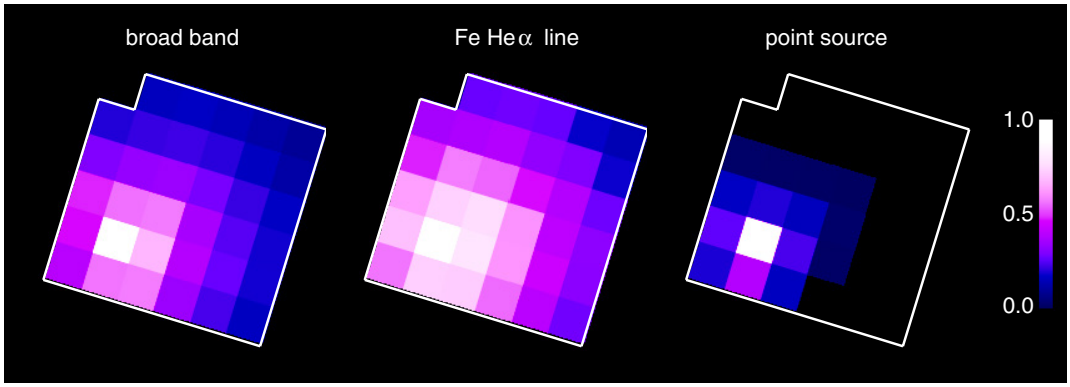
Extended Data Figure 3 | The Fe He-α line complex from the central region around the AGN. The 5.0–8.5-keV spectrum was modelled with an isothermal, optically thin plasma in collisional ionization equilibrium using either APEC/ATOMDB 3.0.3 (red) or SPEX 3.0 (blue), with an additional power-law component accounting for emission from the central AGN. During the fit we excluded the Fe Heα resonance line because this can be affected by resonant scattering of photons by the intracluster gas in the line of sight. The two spectral codes provide similar results with an average temperature of 3.8 ± 0.1 keV and metallicity consistent with the solar value. We obtain a velocity broadening of 156 ± 12 km s^{-1} from APEC and 178 ± 9 km s^{-1} from SPEX. Both models suggest that the resonant line has been suppressed in the central region. Error bars are 1 s.d.

© 2016 Macmillan Publishers Limited. All rights reserved



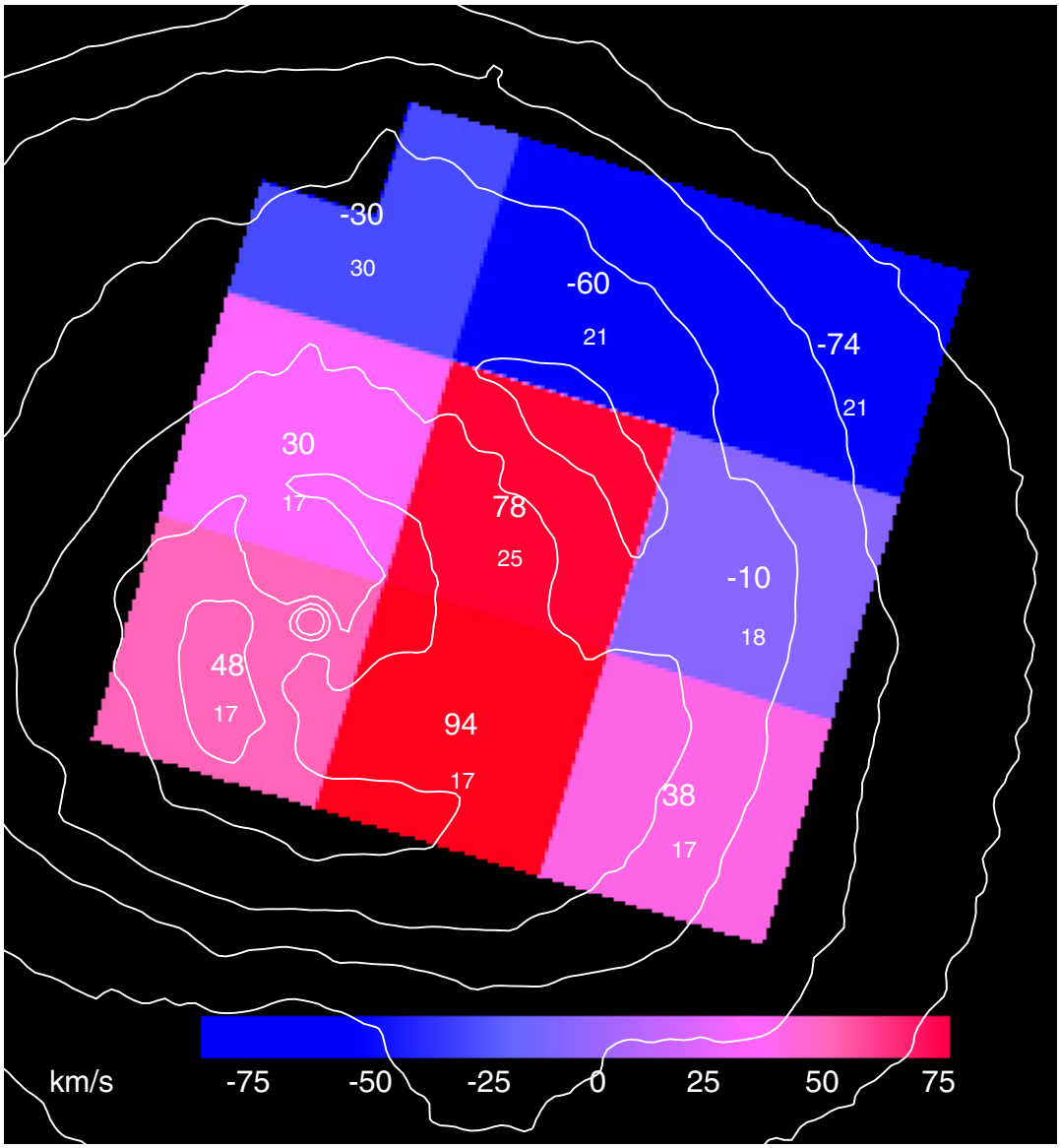
Extended Data Figure 4 | Confidence contours for joint fits of redshift z and velocity broadening σ_v are compared. The three line complexes have been fitted independently. The contours are plotted at $\chi^2_{\rm min}+2.3$ (68%, two parameters) and $\chi^2_{\rm min}+6.17$ (95%). The three fits give consistent redshifts (with the one to which the data were self-calibrated) and broadening.

© 2016 Macmillan Publishers Limited. All rights reserved



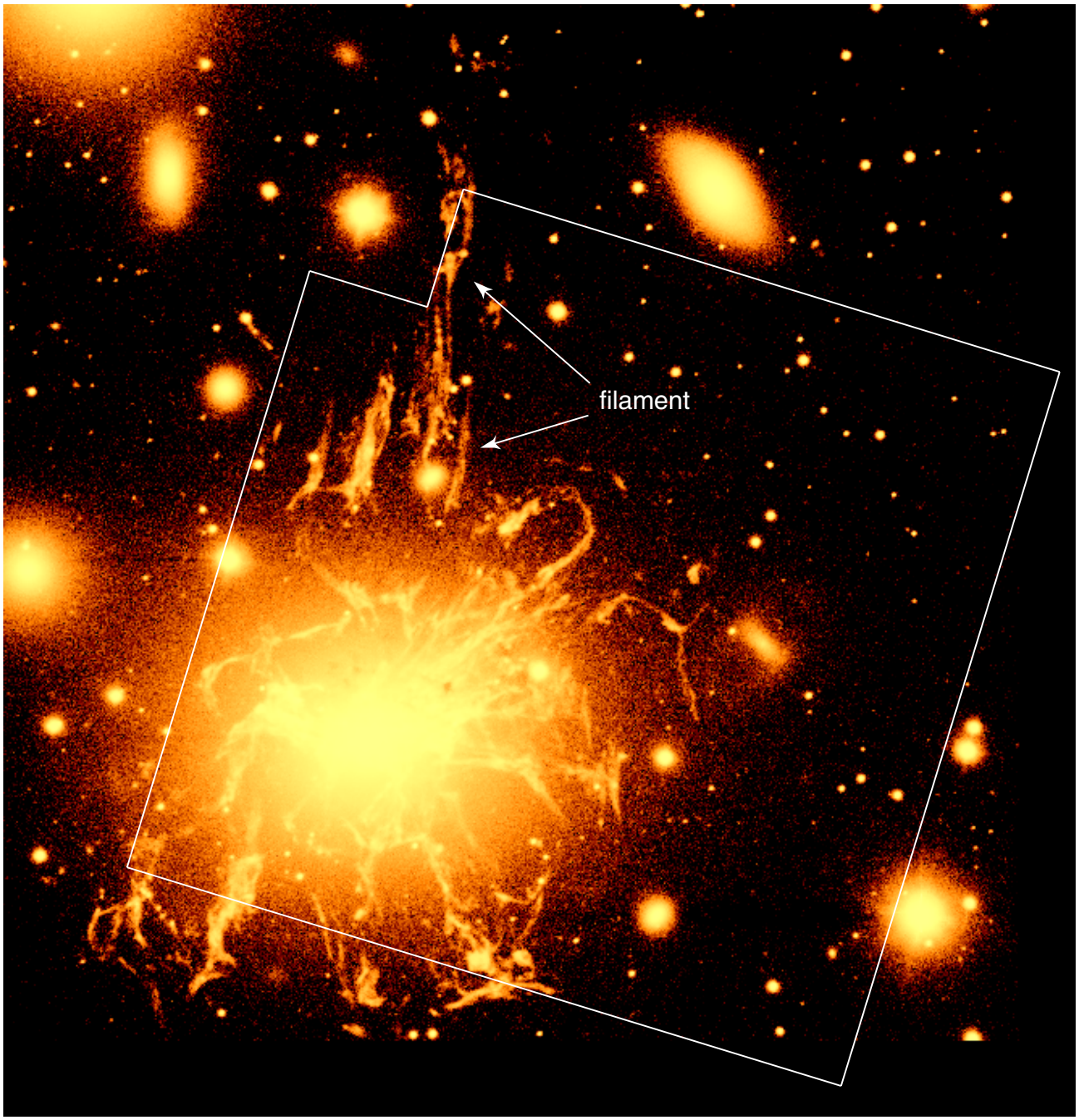
Extended Data Figure 5 | The spatial response of the SXS array. The total broadband counts (colour scale) seen across the detector array (left), Fe Heα line counts (centre) that come mostly from the diffuse cluster plasma, and a model response of a point source centred in the pixel coincident with the nucleus of NGC 1275 (right) are compared. Brightness is normalized to the same peak value.

© 2016 Macmillan Publishers Limited. All rights reserved



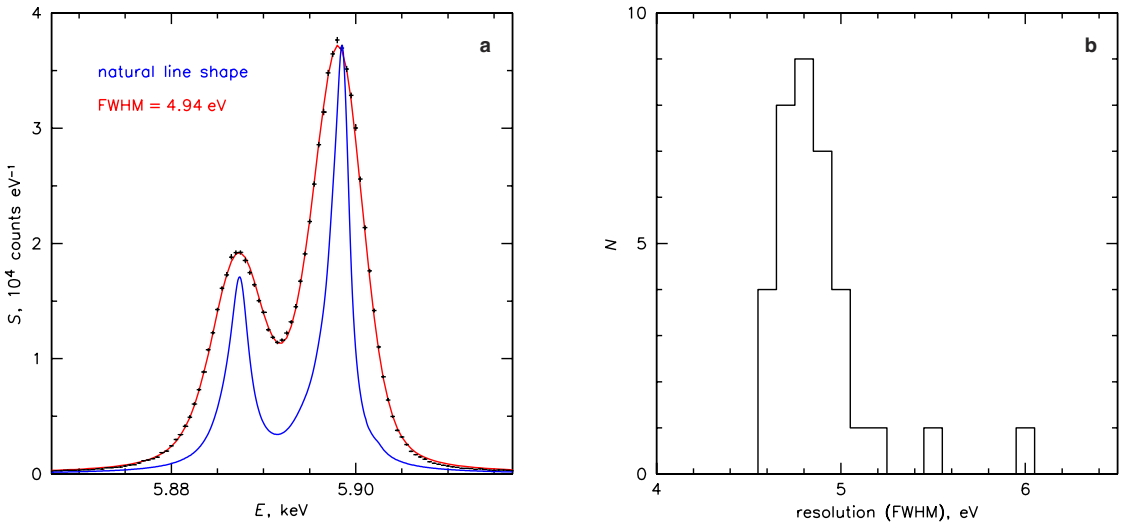
Extended Data Figure 6 | The line-of-sight gas velocities overlaid on a deep Chandra image. The Chandra image is from ref. 32. The contours increase by a factor of 1.5. The numbers in the larger font indicate the velocity in each region (see also the colour scale). The 90% errors in the figure (numbers in the smaller font) are statistical only; our estimate of the calibration uncertainty in individual pixels is 50 km s^{-1}. Heliocentric correction has been applied. Velocities are shown relative to that of NGC 1275, whose redshift is $z = 0.01756$ (ref. 33).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | The SXS field overlaid on the cold gas nebulosity surrounding NGC 1275. The image shows Hα emission[34]. The radial velocity along the long northern filament measured from CO data[21] decreases, south to north (within the SXS field of view), from about $+50\,\mathrm{km\,s^{-1}}$ to $-65\,\mathrm{km\,s^{-1}}$. This is similar to the trend seen in the SXS velocity map (Extended Data Fig. 6).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | In-flight spectral resolution of the SXS. **a**, The composite spectrum of all pixels (excluding the calibration pixel) when they were exposed to the ^{55}Fe source on the filter wheel. The blue line shows the expected natural line shape and the red line shows the observed profile (error bars are 1 s.d.). **b**, A histogram of pixel resolution. *N* is the number of pixels sharing that resolution.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Line energies used in the Gaussian fits

Energy (eV)	λ (Å)	Charge state	Transition	Label	Note	Ref
He-α multiplet						
6617.00	1.8737				Blend – identified in (34) as Be- and Li-like iron	Fit
6628.93	1.8704	XXIII	$1s2s^2 2p\ ^1P_1 \rightarrow 1s^2 2s^2\ ^1S_0$		Be-like	35
6636.84	1.8681	XXV	$1s2s\ ^3S_1 \rightarrow 1s^2\ ^1S_0$	He α (z)	Forbidden	35
6645.24	1.8658	XXIV	$1s2p^2\ ^2D_{5/2} \rightarrow 1s^2 2p\ ^2P_{3/2}$		Li-like	35
6654.19	1.8633	XXIV	$1s2s2p\ ^2P_{1/2} \rightarrow 1s^2 2s\ ^2S_{1/2}$ $1s2p^2\ ^2D_{3/2} \rightarrow 1s^2 2p\ ^2P_{1/2}$		Li-like blend	35
6662.09	1.8610	XXIV	$1s2s2p\ ^2P_{3/2} \rightarrow 1s^2 2s\ ^2S_{1/2}$		Li-like	35
6667.90	1.8594	XXV	$1s2p\ ^3P_1 \rightarrow 1s^2\ ^1S_0$	He α (y)	Intercombination	35
6682.45	1.8554	XXV	$1s2p\ ^3P_2 \rightarrow 1s^2\ ^1S_0$	He α (x)	Intercombination	35
6700.76	1.8503	XXV	$1s2p\ ^1P_1 \rightarrow 1s^2\ ^1S_0$	He α (w)	Resonance	35
H-like doublet						
6951.96	1.7834	XXVI	$2p\ ^2P_{1/2} \rightarrow 1s_{1/2}\ ^2S_{1/2}$	Ly α2		36
6973.18	1.7780	XXVI	$2p\ ^2P_{3/2} \rightarrow 1s_{1/2}\ ^2S_{1/2}$	Ly α1		36
He-β doublet						
7871.31	1.5751	XXV	$1s3p\ ^3P_1 \rightarrow 1s^2\ ^1S_0$	He β2		37
7880.67	1.5733	XXV	$1s3p\ ^1P_1 \rightarrow 1s^2\ ^1S_0$	He β1		37

Data are from refs 34–37.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18314

Photodissociation of ultracold diatomic strontium molecules with quantum state control

M. McDonald[1], B. H. McGuyer[1], F. Apfelbeck[1]†, C.-H. Lee[1], I. Majewska[2], R. Moszynski[2] & T. Zelevinsky[1]

Chemical reactions at ultracold temperatures are expected to be dominated by quantum mechanical effects. Although progress towards ultracold chemistry has been made through atomic photoassociation[1], Feshbach resonances[2] and bimolecular collisions[3], these approaches have been limited by imperfect quantum state selectivity. In particular, attaining complete control of the ground or excited continuum quantum states has remained a challenge. Here we achieve this control using photodissociation, an approach that encodes a wealth of information in the angular distribution of outgoing fragments. By photodissociating ultracold $^{88}Sr_2$ molecules with full control of the low-energy continuum, we access the quantum regime of ultracold chemistry, observing resonant and nonresonant barrier tunnelling, matter–wave interference of reaction products and forbidden reaction pathways. Our results illustrate the failure of the traditional quasiclassical model of photodissociation[4–7] and instead are accurately described by a quantum mechanical model[8,9]. The experimental ability to produce well-defined quantum continuum states at low energies will enable high-precision studies of long-range molecular potentials for which accurate quantum chemistry models are unavailable, and may serve as a source of entangled states and coherent matter waves for a wide range of experiments in quantum optics[10,11].

To obtain full control over the initial (molecular) and final (continuum) quantum states, we photodissociate diatomic strontium molecules ($^{88}Sr_2$) that are optically trapped at a temperature of ~5 μK (ref. 12). These molecules are produced by photoassociating laser-cooled Sr atoms in a far-off-resonant one-dimensional (1D) optical lattice with a depth of up to 50 μK. The Sr atoms are divalent and do not form covalent chemical bonds. However, the ground-state Sr_2 dissociation energy is larger than in typical van der Waals complexes and similar to hydrogen bonded systems such as the water dimer. The $^{88}Sr_2$ molecules that we produce are either weakly bound near the ground state threshold ($^1S + {}^1S$ atomic limit) or, with an extra step of optical preparation, near the lowest singly excited threshold ($^1S + {}^3P_1$). The long-lived (22 μs) excited atomic state $^3P_1$ is responsible for the low laser-cooling temperature, efficient molecule creation, accurate state preparation and high spectroscopic resolution that allows photodissociation very close to the threshold. Photodissociation is driven by a 10–20 μs pulse of linearly polarized 689 nm light (intensity 0.3–30 W cm^{-2}, bandwidth <200 Hz) propagating along the lattice axis. The light frequency is chosen to probe a continuum energy in the range of 0–15 mK because this matches typical electronic and rotational barrier heights. After a controlled delay, the fragments are detected by absorption imaging via the strong $^1S - {}^1P_1$ Sr transition using 461 nm light propagating almost parallel to the lattice axis, so that the initial sample of $>10^4$ molecules appears as a point source. This produces a 2D projection of the 3D spherical shell (Newton sphere) formed by the expanding fragments. The experimental geometry is illustrated in Fig. 1, including the definition of angles θ and ϕ for a dissociating molecule and an image of the fragments showing clear dependence on both angles. The arrangement of the optical lattice, the photodissociating and imaging light, a camera and a small bias magnetic field B that fixes the quantum axis are also shown. In all subsequent images the colour scheme is identical to that of Fig. 1b apart from the overall normalization and the fields of view are 0.1–0.9 mm on each side.

Following photodissociation, the angular distribution of fragment positions is described by an intensity (or differential cross-section)

$$I(\theta,\phi) = |f(\theta,\phi)|^2 \quad (1)$$

which is the square of a scattering amplitude f that can be expanded in terms of partial amplitudes, $f(\theta,\phi) = \sum_{JM} f_{JM}\psi_{JM}(\theta,\phi)$. This expansion uses angular basis functions ψ_{JM} of the outgoing electronic

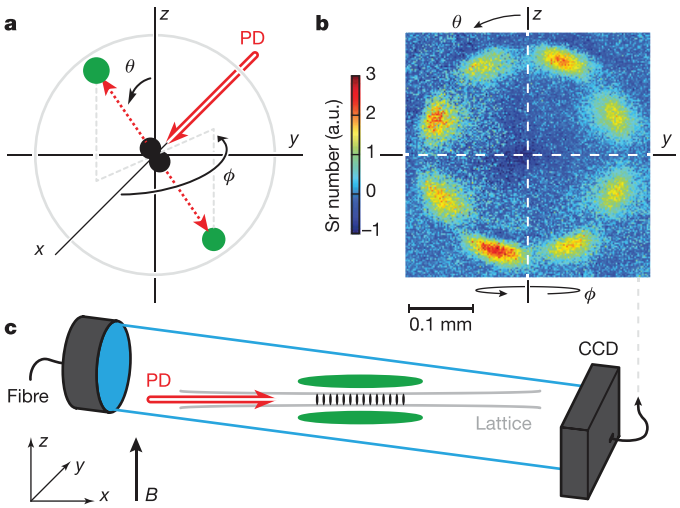


Figure 1 | Photodissociation of diatomic molecules in an optical lattice. **a**, A homonuclear molecule (black circles) producing fragments (green circles) with well-controlled speeds forms a Newton sphere. The distribution of the fragments on the sphere surface is parameterized by a polar angle θ relative to the z axis and an azimuthal angle ϕ relative to the x axis in the xy plane. The photodissociating (PD) light propagates along $+x$. **b**, An experimental image of the fragments corresponds to the Newton sphere projected onto the yz plane. This particular image is one of many we observed that is highly quantum mechanical in nature and distinctly lacks fragments that are emitted along the xz plane. The distribution is thus not cylindrically symmetric about the z axis and depends on ϕ in addition to θ. a.u., arbitrary units. **c**, The fragments (green ovals) are detected by absorption imaging using a charge-coupled device (CCD) camera and a wide light beam from an optical fibre. The photodissociating light is coaligned with the lattice axis along x. The imaging light is nearly coaligned with x (a small tilt is present for technical reasons). A magnetic field can be applied along the z axis.

[1]Department of Physics, Columbia University, 538 West 120th Street, New York, New York 10027-5255, USA. [2]Quantum Chemistry Laboratory, Department of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland. †Present address: Faculty of Physics, Ludwig Maximilian University of Munich, Schellingstrasse 4, 80799 Munich, Germany.

© 2016 Macmillan Publishers Limited. All rights reserved

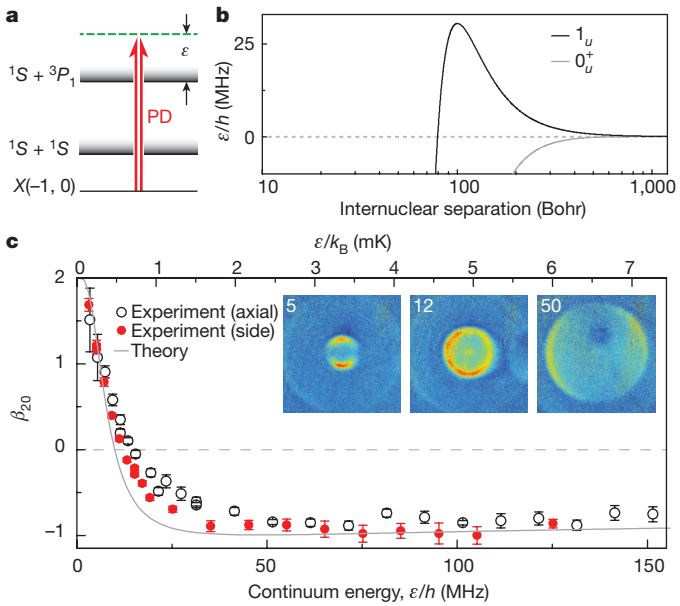


Figure 2 | Photodissociation to a multichannel continuum. a, Schematic for PD of $^{88}Sr_2$ in the initial ground state $X(v_i, J_i)$ to an excited continuum energy ε, which is subsequently expressed in MHz or in mK (via the Boltzmann constant k_B). **b**, Potential energy structure ($\lesssim$1 mK) of the $^1S + {}^3P_1$ continuum, showing both of the electronic potentials (0_u^+ and 1_u) that couple to the ground state via E1 transitions[9]. **c**, The angular anisotropy parameter β_{20} for this process measured by two imaging methods (using axial-view and side-view CCD cameras) and calculated using a quantum chemistry model. The inset images show fragments at three different energies ε/h labelled in MHz. The images and curves indicate a steep change in the angular anisotropy in the 0–2 mK continuum energy range. The experimental errors for axial imaging were estimated by varying the choice of centre point for the pBasex algorithm and averaging the results, and for side imaging from least-squares fitting to equation (2) convolved with a blurring function to account for experimental imperfections.

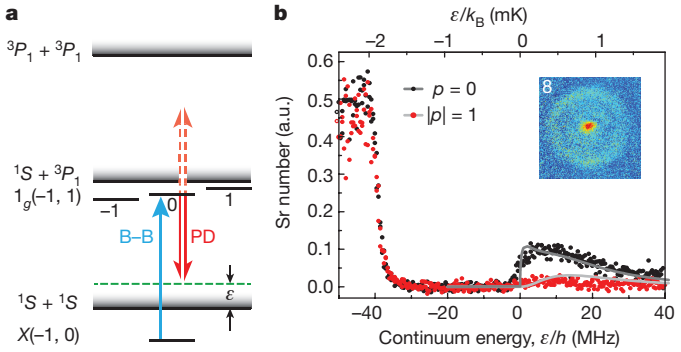


Figure 3 | E1-forbidden photodissociation experiment and theory. a, Molecules in $M_i = 0$ of the long-lived 1_g state below the $^1S + {}^3P_1$ threshold are prepared with a bound–bound (B–B) π pulse and fragmented at the gerade ground continuum with PD light. **b**, M1/E2 photodissociation produces photofragments for $\varepsilon > 0$ (right), and as predicted is strongest for $p = 0$. Solid curves are calculations of the total transition strength using a quantum chemistry model. E1 photodissociation to the $^3P_1 + {}^3P_1$ continuum also appears for $\varepsilon < 0$ (left). The inset image shows fragments for $p = 0$ and $\varepsilon/h \approx 8$ MHz. The strong central dot results from spontaneous E1-forbidden photodissociation of the molecules into low-energy atoms that are captured by the lattice.

channel, where J and M are the total angular momentum and its projection onto the quantum axis, respectively. The intensities for separate electronic channels superpose to produce the total intensity $I(\theta, \phi)$. Cylindrically asymmetric distributions with ϕ dependence are possible if several M states are coherently created because $\psi_{JM}(\theta, \phi) \equiv e^{iM\phi}\psi_{JM}(\theta, 0)$. Our measured angular distributions can be summarized with the parameterization

$$I(\theta, \phi) \propto 1 + \sum_{l=1}^{\infty} \sum_{m=0}^{l} \beta_{lm} \cos(m\phi) P_l^m(\cos\theta) \qquad (2)$$

where $P_l^m(\cos\theta)$ is an associated Legendre polynomial and l is restricted to even values for homonuclear diatomic molecules. The β_{lm} coefficients are directly related to the amplitudes f_{JM}, but hide some of the simplicity that is apparent from using the amplitudes with equation (1). Besides their use in photodissociation, fragment angular distributions are powerful observables in photoionization experiments[13] as they provide a route to completely measure the ionization matrix element amplitudes and phases[14]. The internal angular momenta of the fragments may also carry valuable information[15].

To investigate a multichannel electronic continuum at very low dissociation energies, ε, we prepared ultracold molecules in the $J_i = 0$ initial state of the least-bound vibrational level $v_i = -1$ (negative v_i count down from threshold) of the ground potential X and photodissociated them at the excited $^1S + {}^3P_1$ continuum via the electric dipole (E1) process illustrated in Fig. 2a with an applied field $B = 0$. There are four allowed channels in the excited continuum, which are labelled 0_u^+, 1_u, 0_g^+ and 1_g, where the letters u/g refer to the inversion symmetry of the wave function and the numbers 0/1 refer to the internuclear axis projection of the electronic angular momentum. Only u-symmetric channels are E1-accessible from the ground state. Here, the light polarization sets the quantum axis along z and the fragments can only have $J = 1$, $M = 0$ quantum numbers because $J \geq 1$ for the 0_u^+ and 1_u electronic potentials shown in Fig. 2b. As 1_u has an $\sim$30 MHz ($\sim$1.5 mK) repulsive electronic barrier, we expect the fragment angular distribution to evolve in the probed energy range due to barrier tunnelling. We observe a steep variation of the single anisotropy parameter needed to describe this process, β_{20} from equation (2). Two methods were used to measure this data: axial-view imaging processed with the pBasex algorithm[16] and side-view imaging integrated along the lattice and fitted to a density profile. Figure 2c shows that both methods agree and reveals an evolution of the fragment distribution from a parallel dipole ($\beta_{20} \approx 2$ at $\varepsilon/h \approx 5$ MHz where h is Planck's constant) to a uniform shell ($\beta_{20} \approx 0$ at $\varepsilon/h \approx 12$ MHz) and then a perpendicular dipole ($\beta_{20} \approx -1$ at $\varepsilon/h \approx 50$ MHz). A quantum chemistry model[8,9] was used to calculate the expected anisotropy curve in Fig. 2c by connecting the bound and continuum wave functions via Fermi's golden rule to compute the amplitudes f_{JM}, showing strong qualitative agreement with the data. The theoretical 0_u^+ and 1_u Coriolis-mixed potentials agree well with high-precision bound-state $^{88}Sr_2$ spectroscopy[9,17], but this work is the first test of their predictive power in the continuum.

E1-forbidden photodissociation is an important effect in atmospheric physics and must be considered when calculating the total absorption cross section for molecular oxygen within the so-called Herzberg continuum. Surprisingly, however, neither magnetic dipole (M1) nor electric quadrupole (E2) photodissociation has been directly observed previously. In most cases E1 is also present, making it challenging to study the weaker M1/E2 processes. However, experiments with ultracold Sr_2 allow measurements of pure M1/E2 photodissociation and a comparison with quantum mechanical calculations. Using resonant π pulses we prepare metastable molecules in a $J_i = 1$, $M_i = 0$ state of the least-bound vibrational level of the subradiant 1_g potential that has no E1 coupling to the ground state[12], as sketched in Fig. 3a. The frequency of the dissociating light was varied as shown in Fig. 3b. Here $p = 0$ ($|p| = 1$) implies that the light polarization has a magnetic field parallel (perpendicular) to the quantum axis. The prominent, polarization-independent feature on the left ($\varepsilon < 0$) is E1 photodissociation above the $^3P_1 + {}^3P_1$ threshold, whereas the weaker, polarization-dependent feature on the right is M1/E2 photodissociation. As the figure shows, the strength of this

© 2016 Macmillan Publishers Limited. All rights reserved

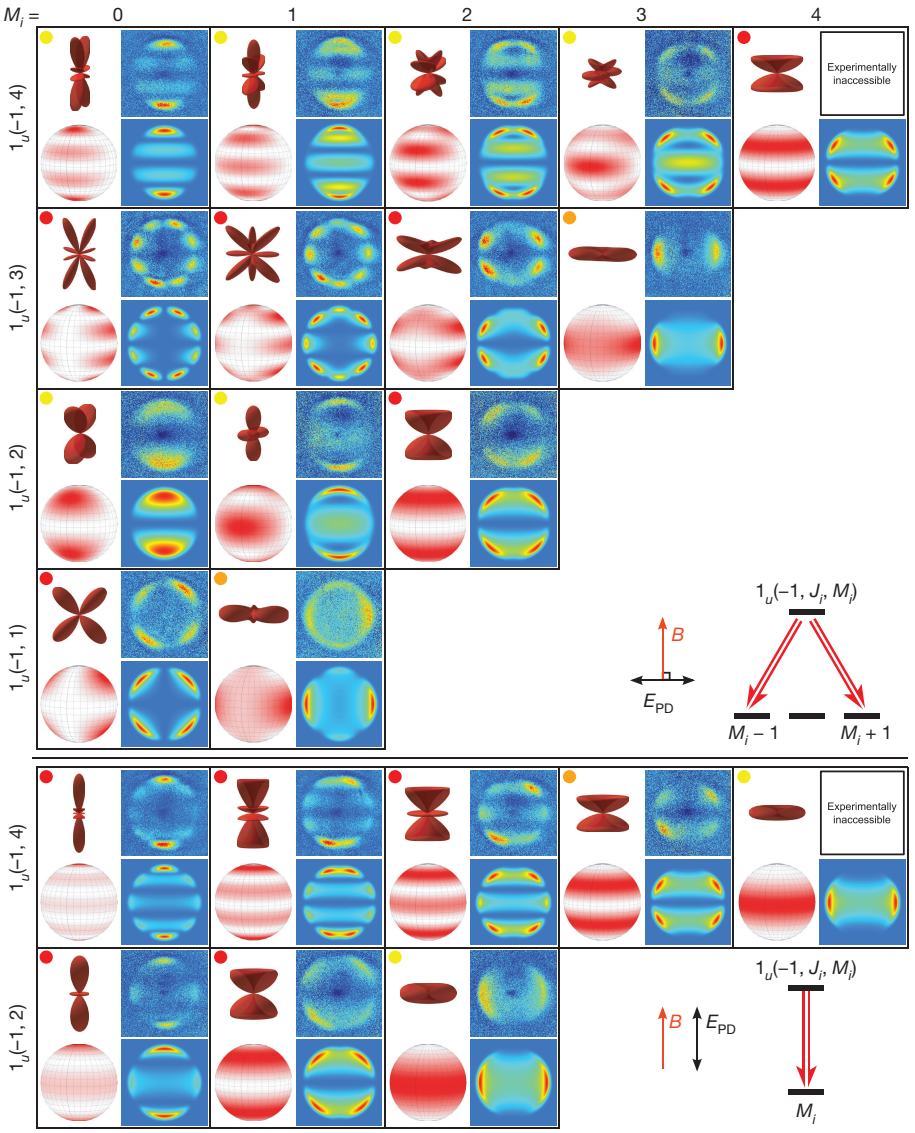


Figure 4 | Photodissociation of singly excited ($^1S + {}^3P_1$) molecules to the ground-state continuum with energies of several millikelvin. Each row and column corresponds to molecules prepared in the indicated $1_u(v_i, J_i)$ state and M_i sublevel. ($M_i = 4$ was not accessible experimentally.) The upper and lower sections correspond to PD light polarizations $|p| = 1$ and 0, respectively, where the PD laser's electric field is E_{PD}. Within each square panel, the experimental image is on the top right, with a comparable simulation of a projected Newton sphere on the bottom right. The full sphere rendition is on the bottom left and the top left shows the mapping of the fragment detection probability at each angle onto the radial coordinate of a surface. For $|p| = 1$, matter–wave interference occurs if two values of M are produced, leading to strongly ϕ-dependent patterns. For each case, the degree of agreement with the quasiclassical approximation is indicated by a coloured dot, as explained in the text.

forbidden process tapers off rapidly and is substantial only below ~1 mK. The inset displays fragments near the peak of the $p = 0$ spectrum. Although the number of fragments for $p = 0$ is unaffected by interference between M1 and E2 pathways, our calculations indicate that their angular distributions (Extended Data Fig. 1) are sensitive to this rarely observed interference.

We take advantage of the single-channel spinless ground state of $^{88}Sr_2$ to explore chemistry in the ultracold regime, obtain a library of fragment distributions and test a quasiclassical model of photodissociation. We prepare singly excited ($^1S + {}^3P_1$) molecules with quantum numbers J_i, M_i and immediately photodissociate them at the $^1S + {}^1S$ ground state continuum, in some cases applying B up to 20 G to enable symmetry-forbidden E1 transitions[18]. To control the final value of J in the continuum (which quantum statistics requires to be even for bosonic ground-state $^{88}Sr_2$) we either obtain a unique J by choosing to start from an even J_i and taking advantage of selection rules, or, if multiple 'partial waves' with different J are possible and interfere, we choose an ε value at which a single J wave strongly dominates, as discussed below. To control the final values of M we orient the linear polarization of the photodissociating light either parallel ($p = 0$) to the quantum axis, for which selection rules ensure $M = M_i$, or perpendicular ($|p| = 1$), for which $M = M_i \pm 1$. Thus, we are able to engineer and image different continua in either pure M states or as their coherent quantum interference. Disruption from Zeeman shifts is avoided because the ground continuum is practically nonmagnetic.

Figure 4 shows a full range of distributions parameterized by equation (1) with either $f(\theta, \phi) = Y_{J,M_i}(\theta, \phi)$ or $f(\theta, \phi) = \sqrt{R}\, Y_{J,M_i-1}(\theta, \phi) + e^{i\delta}\sqrt{1-R}\, Y_{J,M_i+1}(\theta, \phi)$. Here the spherical harmonics $Y_{JM} = \Psi_{JM}$ for the ground continuum, R and δ are the relative amplitude and phase parameters and $J = 2$ or 4. (At the chosen continuum energies, the $p = 0$ patterns for $J_i = 1, 3$ would be nearly redundant with $J_i = 2, 4$ and so are omitted.) Quantum mechanical calculations, included for comparison, assume that the continuum states are dominated by the higher J contribution. Figure 4 suggests the following observations. First, the coherent superposition of a pair of M, which occurs for $|p| = 1$ but not $p = 0$, leads to clean observations of distributions without

© 2016 Macmillan Publishers Limited. All rights reserved

cylindrical symmetry, previously unreported for diatomic molecules. In particular, multiple cases are shown of a molecule fragmenting into up to eight distinct (θ, ϕ) regions. Second, the same final states ($J=4$, $M=\pm1$) are produced for $|p|=1$, $M_i=0$ and $J_i=4$, 3 at the chosen continuum energies. Thus we could expect to observe identical fragment patterns. However, a subtle point is that odd J_i and even J_i produce $M=M_i\pm1$ probability amplitudes with an opposite relative phase. This results in identical ϕ-dependent patterns rotated by 90° relative to each other. The same mapping of the relative phase onto the rotation angle occurs for $|p|=1$, $M_i=0$ and $J_i=2$, 1. Third, the previous point roughly holds for the higher values of M_i as well, but non-identical populations of $M=M_i\pm1$ are produced due to asymmetrical coupling strengths. For example, the matter–wave interference patterns for $(J_i, M_i)=(4, 2)$ and $(3, 2)$ are not only rotated relative to each other, but have slightly different shapes.

Over the past few decades a quasiclassical model has been advanced to predict the angular distributions for single-photon E1 photodissociation of diatomic molecules prepared in arbitrary quantum states[6,7,19]. This approach multiplies the conventional distribution[4,5] for molecules prepared in spherically symmetric states or ensembles, $I(\theta)\propto 1+\beta_{20}P_2^0(\cos\chi)$, by a probability density $|\Phi_i|^2$ for the initial molecular axis orientation, which gives $I(\theta,\phi)\propto|\Phi_i(\theta,\phi)|^2[1+\beta_{20}P_2^0(\cos\chi)]$ where $\chi=(\theta,\phi)$ is the polar angle defined with respect to the orientation of linear polarization of the photodissociating light and (θ,ϕ) are defined by the quantization axis, as before. This intuitive model suggests that photodissociation probes the 'shape' of the initial molecules, as detailed in Extended Data Fig. 2. Its validity, however, has been questioned over the years[20].

To indicate the level of agreement with the quasiclassical model, we include coloured dots for each pattern in Fig. 4. A green dot indicates exact agreement between the quasiclassical and quantum mechanical calculations, a yellow dot indicates qualitative agreement that cannot be made exact by adjusting β_{20}, an orange dot indicates disagreement that can become a qualitative agreement by adjusting β_{20} and a red dot indicates clear disagreement for all β_{20}—usually because fragments are observed where $|\Phi_i|^2$ has a node. For all cases in Fig. 4 the quasiclassical model fails to varying degrees. Although this could be expected for the 1_u initial states[19], surprisingly even photodissociation of the 0_u^+ states (Extended Data Fig. 3) disagrees with the quasiclassical model in all cases where more than a single J is possible in the continuum. This is because only the single-J cases allow the quasiclassical assumption of prompt axial recoil to be satisfied at such low continuum energies. Furthermore, our experiments demonstrate that initial molecules with different shapes (for example, 0_u^+ versus 1_u) can produce nearly identical distributions, highlighting that the fragment distributions are solely determined by the final (continuum) states.

Ultracold photodissociation readily reveals features of the continuum just above the threshold. The ability to freely explore a large range of continuum energies, together with strict optical selection rules and cleanly prepared quantum states, provides a versatile tool to isolate and study individual reaction channels. Whereas Fig. 2 explored tunnelling through an electronic barrier, Fig. 5 shows the case when only rotational barriers are present. Here molecules prepared in the $0_u^+(v=-3, J_i=3, M_i=0)$ state are photodissociated with $p=0$, resulting in continuum states with $M=0$ and $J=2$, 4. This mixture can be described by equation (1) with $f(\theta,\phi)=\sqrt{R}\,Y_{20}(\theta,\phi)+e^{i\delta}\sqrt{1-R}\,Y_{40}(\theta,\phi)$. Figure 5a is a plot of the branching ratio R and the interference amplitude $2\cos\delta\sqrt{R(1-R)}$ for the 0–15 mK range of continuum energies. The data show a good qualitative agreement with quantum chemistry calculations, and reveal a predicted but so far unobserved g-wave shape resonance (or quasi-bound state) confined by the $J=4$ centrifugal barrier. This long-lived (~10 ns) resonance ~66 MHz above threshold could be used to control light-assisted molecule formation rates[21]. Shape resonances can also be mapped with magnetic Feshbach dissociation of ground-state molecules[22–24]. However, photodissociation is more widely applicable to molecules with any type of spin structure in any electronic state, and allows more control over the quantum numbers. In Fig. 5b an anisotropic, energy-independent pattern is visible on all images with a radius close to that of the 62 MHz image. We have confirmed that this signal arises from spontaneous photodissociation of the molecules into the g-wave shape resonance (Extended Data Fig. 4).

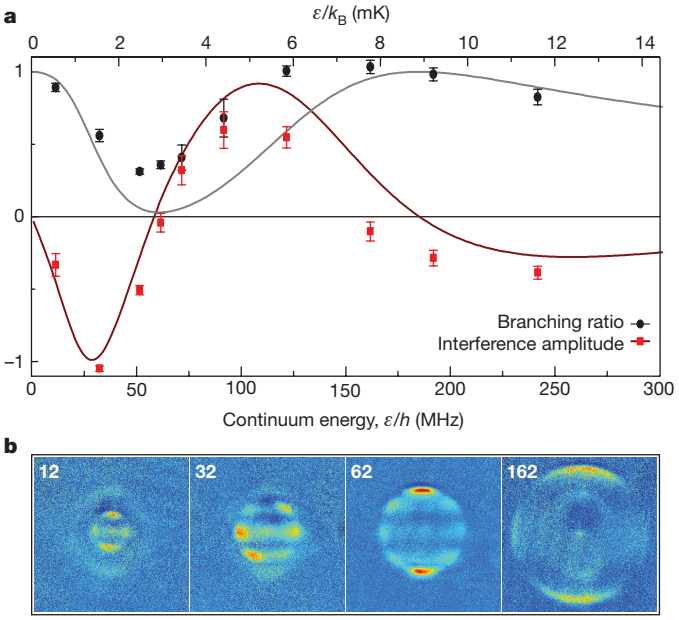


Figure 5 | Energy-dependent photodissociation near a shape resonance. **a**, Molecules prepared in the $0_u^+(v_i=-3, J_i=3, M_i=0)$ state are photodissociated at the ground continuum. For $p=0$, selection rules lead to a single $M=0$ but a mixture of $J=2$, 4. The branching ratio and interference amplitude of this mixture, as described in the text, evolve with energy and reveal a $J=4$ (g-wave) shape resonance at ~3 mK. The experimental data were analysed with pBasex and errors were estimated by varying the effective saturation intensity, used to process the absorption images, within its uncertainty. The theoretical curves were calculated with a quantum chemistry model. **b**, Images of fragments labelled by their continuum energies ε/h in MHz that show the evolution with energy. The faint anisotropic, energy-independent pattern with roughly the same radius as the 62 MHz image is from spontaneous decay into the shape resonance.

This work explores light-induced molecular fragmentation in the fully quantum regime. Quasiclassical descriptions are not applicable and our observations are dominated by coherent superpositions of matter waves originating from monoenergetic continuum states with different quantum numbers. The results agree with a state-of-the-art quantum chemistry model[8,9], but challenge the theory to describe more complicated phenomena. For example, preliminary observations of photodissociation to the doubly excited continuum (as in Fig. 3b) indicate rich structure near the threshold. This continuum is not well understood, while interactions near the $^3P_1+{}^3P_1$ threshold play a key role in recent proposals and experiments in ultracold many-body science[25]. Other excited continua with even longer lifetimes (for example, the subradiant 1_g and 0_g^+ manifolds) exist for Sr_2 and similar molecules and should enable the exploration of entangled continuum states. Photodissociation can shed light on the ultracold chemistry of a rich array of molecular states, as well as on new reaction mechanisms—as was shown here with M1/E2 photodissociation. With improved control of the imaging and of the optical lattice effects, experiments can get even closer to the threshold. We expect to reach nanokelvin fragment energies in the lattice, leading to high-precision measurements of binding energies for tests of fundamental physics and molecular quantum electrodynamics[26,27]. Ultralow fragment energies can also aid in the creation of novel ultracold atomic gases[28]. A promising future direction would be to enhance the quantum control achieved

© 2016 Macmillan Publishers Limited. All rights reserved

here by manipulating the final continuum states with external fields[29,30]. We have shown the extreme sensitivity of weakly bound molecules to small magnetic fields[18], and the same principle applies just above threshold. This external control over ultracold chemistry should allow the study and manipulation of new reaction pathways.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 January; accepted 26 April 2016.

1. Jones, K. M., Tiesinga, E., Lett, P. D. & Julienne, P. S. Ultracold photoassociation spectroscopy: long-range molecules and atomic scattering. *Rev. Mod. Phys.* **78,** 483–535 (2006).
2. Chin, C., Grimm, R., Julienne, P. & Tiesinga, E. Feshbach resonances in ultracold gases. *Rev. Mod. Phys.* **82,** 1225–1286 (2010).
3. Ospelkaus, S. *et al.* Quantum-state controlled chemical reactions of ultracold potassium-rubidium molecules. *Science* **327,** 853–857 (2010).
4. Zare, R. N. & Herschbach, D. R. Doppler line shape of atomic fluorescence excited by molecular photodissociation. *Proc. IEEE* **51,** 173–182 (1963).
5. Zare, R. N. Photoejection dynamics. *Mol. Photochem.* **4,** 1–37 (1972).
6. Choi, S. E. & Bernstein, R. B. Theory of oriented symmetric-top molecule beams: precession, degree of orientation, and photofragmentation of rotationally state-selected molecules. *J. Chem. Phys.* **85,** 150–161 (1986).
7. Zare, R. N. Photofragment angular distributions from oriented symmetric-top precursor molecules. *Chem. Phys. Lett.* **156,** 1–6 (1989).
8. Skomorowski, W., Pawłowski, F., Koch, C. P. & Moszynski, R. Rovibrational dynamics of the strontium molecule in the $A\,^1\Sigma_u^+$, c $^3\Pi_u$, and $a\,^3\Sigma_u^+$ manifold from state-of-the-art ab initio calculations. *J. Chem. Phys.* **136,** 194306 (2012).
9. Borkowski, M. *et al.* Mass scaling and nonadiabatic effects in photoassociation spectroscopy of ultracold strontium atoms. *Phys. Rev. A* **90,** 032713 (2014).
10. Grangier, P., Aspect, A. & Vigue, J. Quantum interference effect for two atoms radiating a single photon. *Phys. Rev. Lett.* **54,** 418–421 (1985).
11. Kheruntsyan, K. V., Olsen, M. K. & Drummond, P. D. Einstein–Podolsky–Rosen correlations via dissociation of a molecular Bose–Einstein condensate. *Phys. Rev. Lett.* **95,** 150405 (2005).
12. McGuyer, B. H. *et al.* Precise study of asymptotic physics with subradiant ultracold molecules. *Nature Phys.* **11,** 32–36 (2015).
13. Reid, K. L. Photoelectron angular distributions. *Annu. Rev. Phys. Chem.* **54,** 397–424 (2003).
14. Hockett, P., Wollenhaupt, M., Lux, C. & Baumert, T. Complete photoionization experiments via ultrafast coherent control with polarization multiplexing. *Phys. Rev. Lett.* **112,** 223001 (2014).
15. Rakitzis, T. P., Kandel, S. A., Alexander, A. J., Kim, Z. H. & Zare, R. N. Photofragment helicity caused by matter–wave interference from multiple dissociative states. *Science* **281,** 1346–1349 (1998).
16. Garcia, G. A., Nahon, L. & Powis, I. Two-dimensional charged particle image inversion using a polar basis function expansion. *Rev. Sci. Instrum.* **75,** 4989–4996 (2004).
17. McGuyer, B. H. *et al.* High-precision spectroscopy of ultracold molecules in an optical lattice. *New J. Phys.* **17,** 055004 (2015).
18. McGuyer, B. H. *et al.* Control of optical transitions with magnetic fields in weakly bound molecules. *Phys. Rev. Lett.* **115,** 053001 (2015).
19. Beswick, J. A. & Zare, R. N. On the quantum and quasiclassical angular distributions of photofragments. *J. Chem. Phys.* **129,** 164315 (2008).
20. Seideman, T. The analysis of magnetic-state-selected angular distributions: a quantum mechanical form and an asymptotic approximation. *Chem. Phys. Lett.* **253,** 279–285 (1996).
21. González-Férez, R. & Koch, C. P. Enhancing photoassociation rates by nonresonant-light control of shape resonances. *Phys. Rev. A* **86,** 063420 (2012).
22. Volz, T. *et al.* Feshbach spectroscopy of a shape resonance. *Phys. Rev. A* **72,** 010704(R) (2005).
23. Mark, M. *et al.* Stückelberg interferometry with ultracold molecules. *Phys. Rev. Lett.* **99,** 113201 (2007).
24. Knoop, S. *et al.* Metastable Feshbach molecules in high rotational states. *Phys. Rev. Lett.* **100,** 083002 (2008).
25. Zhang, X. *et al.* Spectroscopic observation of SU(*N*)-symmetric interactions in Sr orbital magnetism. *Science* **345,** 1467–1473 (2014).
26. Bartenstein, M. *et al.* Precise determination of ^{6}Li cold collision parameters by radio-frequency spectroscopy on weakly bound molecules. *Phys. Rev. Lett.* **94,** 103201 (2005).
27. Salumbides, E. J. *et al.* Bounds on fifth forces from precision measurements on molecules. *Phys. Rev. D* **87,** 112008 (2013).
28. Lane, I. C. Production of ultracold hydrogen and deuterium via Doppler-cooled Feshbach molecules. *Phys. Rev. A* **92,** 022511 (2015).
29. Lemeshko, M., Krems, R. V., Doyle, J. M. & Kais, S. Manipulation of molecules with electromagnetic fields. *Mol. Phys.* **111,** 1648–1682 (2013).
30. Stapelfeldt, H. & Seideman, T. Colloquium: Aligning molecules with strong laser pulses. *Rev. Mod. Phys.* **75,** 543–557 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We gratefully acknowledge ONR grant N000-14-14-1-0802, NIST award 60NANB13D163, and NSF grant PHY-1349725 for partial support of this work, and thank A. T. Grier, G. Z. Iwata and M. G. Tarallo for discussions. R.M. acknowledges the Foundation for Polish Science for support through the MISTRZ programme.

Author Contributions M.M., B.H.M., F.A., C.-H.L. and T.Z. designed the experiments, carried out the measurements and interpreted the data. I.M. and R.M. carried out the calculations and interpreted the data. All authors contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.Z. (tz@phys.columbia.edu).

Reviewer Information *Nature* thanks D. Chandler and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Experimental details. After laser cooling a gas of atomic Sr in a 1D optical lattice, molecules were created via photoassociation to the $0_u^+(v=-4, J=1)$ excited state (binding energy 1,084 MHz) followed by well-directed spontaneous emission to the $X^1\Sigma_g^+(v=-1)$ ground states with $J=0$ or 2 (binding energies of 137 and 67 MHz, respectively)[17,31]. Any remaining atoms were removed with a pulse of imaging light. The molecular sample trapped in the lattice is about 20 μm in diameter and 200 μm long. To prepare metastable $1_g(v_i=-1, J_i=1)$ excited states (binding energy 19 MHz, lifetime ~5 ms), we used a lattice wavelength of ~910 nm to enable resonant 689 nm π pulses to transfer the population from $X(v=-1, J=0)$ to this state before photodissociation[12]. For our experimental conditions, this transfer was ~40% efficient. To prepare shorter-lived 0_u^+ or 1_u excited states, we used a 689 nm light pulse to drive a resonant bound–bound transition from either the $J=0$ or 2 ground state to the desired state during photodissociation. In both cases, we used the polarization of this light and excited-state Zeeman shifts[12,18,32] to select M_i. For reference, the binding energies for the $1_u(v_i=-1, J_i)$ excited states are 353 MHz for $J'=1$; 287 MHz for $J'=2$; 171 MHz for $J'=3$ and 56 MHz for $J'=4$; for the $0_u^+(v_i=-3, J_i=3)$ state, the binding energy is 132 MHz and for $0_u^+(-4,1)$ it is 1,084 MHz. The $^1S+{}^1S$ and $^1S+{}^3P_1$ thresholds may be spectroscopically located with kilohertz precision using the lineshape model of ref. 17.

The photodissociating light propagates along the tight-confinement x axis of the optical lattice (Gaussian waist ~40 μm), and is linearly polarized along either the y axis or the z axis. Except for Fig. 2, for which the net magnetic field is nearly zero, a field of a few to a few tens of gauss is applied along the z axis to fix a quantization axis for excited bound states. The ground bound and continuum states are insensitive to this field, so to avoid mixed-quantization effects from tensor light shifts[18] the optical lattice was linearly polarized along the z axis. We confirmed that our results are unaffected by the small lattice trap depth (typically 0.6–0.8 MHz). A full description of the laboratory-frame spherical tensor components of the fields driving the photodissociation transitions is available in the Supplementary Information.

After the photodissociating light pulse, the fragments were allowed to expand kinetically for several hundred microseconds before their positions were recorded with standard absorption imaging[33]. This expansion time is needed to mitigate blurring due to the finite pulse width and limited imaging resolution, but has the cost of diluting the signal over a larger area, which makes imaging artefacts more problematic. Therefore, we adjusted this expansion time as needed to optimize the signal-to-noise ratio and angular resolution.

Most absorption images were taken with imaging light aligned nearly along the x axis, projecting the fragment positions into the yz plane. Several hundred absorption images were averaged to produce a final record of the fragment positions. To remove imaging artefacts and incidental absorption from unwanted atoms, the experimental sequence was alternated so that every other image contained none of the desired fragments, but everything else. The final image was then computed as the averaged difference between these interlaced 'with fragment' and 'without fragment' images. For Fig. 2 insets and side-view data, we also used an optical pulse to deplete the ground-state population with $J=2$ before photodissociating the $J_i=0$ states.

Forbidden photodissociation angular distributions. A comparison of experimental images of fragment distributions and calculations for the M1/E2 photodissociation of Fig. 3 is presented in Extended Data Fig. 1. Note that a large light intensity was required to drive the forbidden photodissociation process sufficiently rapidly to observe these angular distributions. Besides power broadening the line shapes in Fig. 3b, this high intensity may have affected the measured fragment distributions in Extended Data Fig. 1.

Quasiclassical model. In the photodissociation literature there is a well-known quasiclassical model describing the angular distribution of fragments produced by the photodissociation of aligned molecules

$$I_{qc}(\theta,\phi) = |\Phi_i(\theta,\phi)|^2[1+\beta_{20}P_2^0(\cos\chi)] \quad (3)$$

where the angles are defined in the main text. For homonuclear diatomic molecules in the Born–Oppenheimer approximation, the probability density for the internuclear axis orientation of an initial state with quantum numbers J_i, M_i and $|\Omega_i|$ is given by Wigner–D functions as

$$|\Phi_i(\theta,\phi=0)|^2 = \frac{(2J_i+1)}{8\pi}(|D^{J_i}_{M_i\Omega_i}(0,\theta,0)|^2 + |D^{J_i}_{M_i,-\Omega_i}(0,\theta,0)|^2) \quad (4)$$

where Ω is the internuclear projection of the electronic angular momentum.

We observe disagreement with the quasiclassical model in the majority of cases. At first glance this is surprising because, theoretically, the quasiclassical model has been shown to be either equivalent or a good approximation to the quantum mechanical result for most cases of one-photon E1 photodissociation of a diatomic molecule with prompt axial recoil[19]. However, our measurements are performed at very low continuum energies to reach the ultracold chemistry regime, and thus may violate the assumption of axial recoil[5]. Additionally, ref. 19 predicted that the quasiclassical model should fail for the special case of 'perpendicular' transitions ($|\Delta\Omega|=1$) with initial states that are a superposition of Ω_i states differing by ±2. This special case includes our measurements of 1_u initial states in Fig. 4, and our observations support this prediction.

Extended Data Fig. 2 compares the quasiclassical model with both quantum mechanical predictions and experimental images for several cases. For each, the construction of the quasiclassical prediction is outlined. As in Fig. 4, we use coloured dots to indicate the level of agreement between the two predictions. To determine this agreement, the quasiclassical model neglected the nonadiabatic Coriolis mixing of $|\Omega_i|$ (ref. 32). There is also some ambiguity in choosing a value of β_{20} to use with the quasiclassical model. Conventionally, β_{20} should be equal to 2 for parallel transitions with $\Delta\Omega=0$ and to −1 for perpendicular transitions with $|\Delta\Omega|=1$. In cases of persistent disagreement, we varied β_{20} as a free parameter within the physically allowed range of [−2, 1]. Such a variation has been considered previously as an effect of the breakdown of the axial-recoil approximation[34].

We do observe three cases of exact agreement (indicated by green dots in Extended Data Fig. 3), two of which are highlighted in Extended Data Fig. 2. The reason the quasiclassical model gives exact results is that selection rules only allow a single J in these cases, making the axial-recoil approximation no longer necessary. Specifically, these cases correspond to 0_u^+ initial states with odd J_i for either $|M_i|=J_i$ with $p=0$ or $J_i=1$ and $M_i=0$ with $|p|=1$, for which the angular distribution is energy independent. Agreement occurred here without needing to adjust β_{20}.

In Fig. 2, $p=0$ and the initial state $J_i=M_i=0$ is spherically symmetric, so the angular distribution is parameterized only by β_{20}. Thus, the quasiclassical model can always be adjusted to agree at any continuum energy.

Photodissociation of 0_u^+ states. Single-photon E1 photodissociation of 0_u^+ excited states to the ground continuum is shown in Extended Data Fig. 3, in analogy with Fig. 4 for 1_u states. In Fig. 4 and Extended Data Figs 2–4, the sign of M_i does not affect the results, and our experiments used $M_i>0$ for some of the data sets and $M_i<0$ for others. To avoid confusion we did not label the figures with $|M_i|$, which suggests a superposition of M_i, but instead chose M_i to be positive in the figures.

Spontaneous photodissociation. Extended Data Fig. 4a contains images of the fragments following spontaneous decay of the excited state $0_u^+(v_i=-3, J_i=3, M_i)$ to the ground continuum. As we selectively populate individual M_i sublevels, the measured distributions are anisotropic. They are well described by the incoherent superposition

$$I(\theta) \approx \sum_M |Y_{4M}(\theta,\phi=0)|^2 \begin{pmatrix} 4 & 1 & 3 \\ -M & M-M_i & M_i \end{pmatrix}^2 \quad (5)$$

Here, J is restricted to 4 because the strongest decay is to the $J=4$ shape resonance in the ground continuum. If all M_i were equally populated, which would add a sum over M_i to equation (5), then the distribution would be isotropic.

The shape resonance aids the measurement of the angular distributions because it favours a narrow range of continuum energies. Extended Data Figure 4b contains the results of pBasex analysis of the inset image and highlights how the radial distribution of the atomic fragments is clustered around 66 MHz, revealing an ~10 ns shape resonance lifetime. Extended Data Figure 4c shows that the angular distribution from this analysis matches expectations from equation (5).

Absorption images in figures. Supplementary Tables 1–3 list the parameters used to generate the theoretical images shown in Fig. 4 and Extended Data Figs 1–3. To display theoretical results as simulated absorption images, the intensities are projected into the yz plane by integrating over the x direction. To approximate the blurring present in experimental images from limited optical resolution and light pulse durations, the image is convolved with a Gaussian distribution

$$N(y,z) \propto \int_{-\infty}^{\infty} e^{-\left(\sqrt{x^2+y^2+z^2}-R_0\right)^2/(2\sigma^2)} I(\theta,\phi)\,dx \quad (6)$$

where R_0 is the mean radius, σ is the standard deviation, $\theta=\cos^{-1}\left(z/\sqrt{x^2+y^2+z^2}\right)$ and $\phi=\sin^{-1}\left(y/\sqrt{x^2+y^2}\right)$. The fractional blur was $\sigma/R_0=0.05$ except for Extended Data Fig. 4, where $\sigma/R_0=0.2$.

The same colouring scheme (Matlab colourmap jet) is used in all experimental and theoretical absorption images, up to differences between CMYK and RGB colour mode presentation. Each image was linearly rescaled to fit the finite range [0, 1] of this scheme. To ensure that the same colour corresponds to zero absorption in all images, despite the presence of noise and imaging artefacts, the experimental images are scaled to have an average value of 0.25 in zero-absorption regions and a maximum value of 1. Likewise, the theoretical images are scaled to have

© 2016 Macmillan Publishers Limited. All rights reserved

a minimum value of 0.25 but a maximum value of 0.85 instead of 1, to be more visually similar to experimental images.

The field of view differs between experimental images because of cropping for presentation, and falls in the range of 0.1–0.9 mm on each side. For a given image, the field of view may be accurately determined by calculating the maximum diameter D of the photodissociation products as $D = C\tau\sqrt{(\varepsilon - U)/h}$. Here, the kinetic expansion time τ was 0.3 ms for Fig. 1, 0.8 ms for Fig. 2, 0.6 ms for Fig. 3, 0.3–0.4 ms for Fig. 4, 0.39 ms for Fig. 5, 0.6 ms for Extended Data Fig. 1, 0.3–0.4 ms for Extended Data Fig. 3 and 0.1 ms for Extended Data Fig. 4. The dissociation energies ε not labelled in insets are listed in Supplementary Tables 1–3. From conservation of energy, the parameter $C = 2\sqrt{h/m_{Sr}} \approx 1.348 \times 10^{-4}$ ms$^{-1/2}$ and the lattice depth U must be included as a small offset[17,35]. For Fig. 1b, for example, this gives $D = 0.34$ mm. For theoretical images, D was set to 80% of the image width.

Extracting angular distribution parameters. For angular distributions that are cylindrically symmetric (depend only on θ), the polar basis set expansion (pBasex) algorithm[16] can extract the 3D distribution from 2D projections such as absorption images by fitting the data with the Abel transform of a weighted sum of the Legendre polynomials. We used the software implementation of the pBasex algorithm in ref. 36 to analyse the images in Figs 2 and 5 and Extended Data Fig. 4. For low signal-to-noise images, we found that the extracted distribution is artificially skewed towards spherical symmetry[35]. To eliminate this systematic error, we performed pBasex inversion on a background image made from the set of without-fragment images that is processed to remove imaging artefacts and rescaled so that the average value equals that of the background regions in the final image. The final distribution is then the difference between those extracted for the original image and for the background image. The parameters β_{20} of Fig. 2 and R and δ of Fig. 5 were determined from least-squares fitting of the number of fragments versus θ in the final distribution.

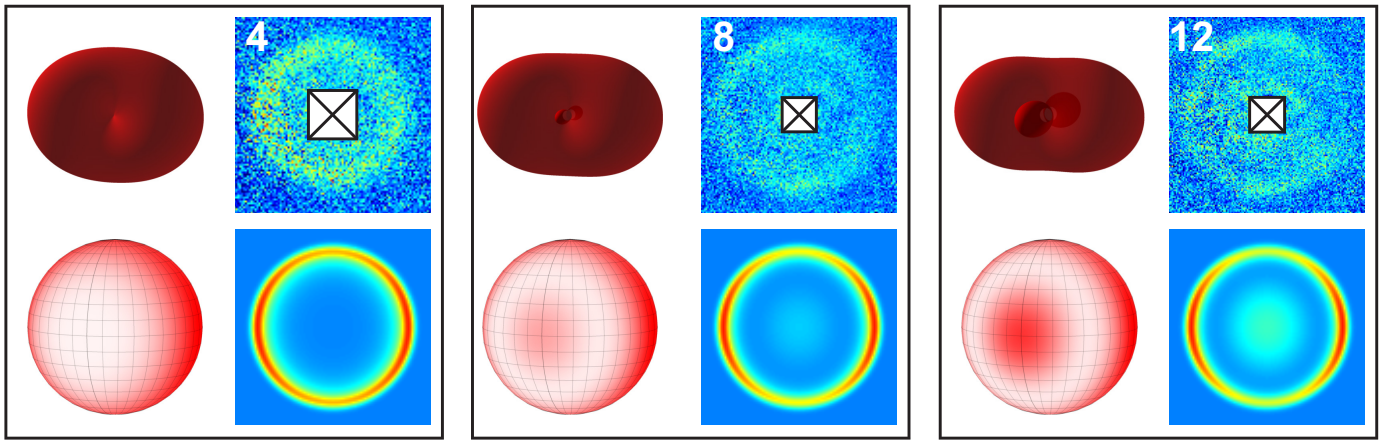
In some cases, such as with the $\varepsilon/h = 32$ MHz inset of Fig. 5b, experimental issues may lead to images with deviations from the expected cylindrical symmetry. This may occur, for example, from imperfect control of the photodissociating light polarization, which may introduce a 'skewness' in the distribution. Apparent deviations from perfect cylindrical symmetry may also have occurred because of absorption imaging error induced by imperfectly correcting for the atomic saturation, which is especially important when the imaging beam exhibits substantial variations across its spatial profile (as was the case for our experiment). In such cases, we proceeded with pBasex analysis but included an estimate of the resulting bias when determining error bars.

For Fig. 2, further analysis was performed by integrating 2D projections along y to convert the images to 1D curves along z. This allows parameters such as β_{20} to be directly extracted by fitting the 1D curve with the expected angular distribution, similar to Extended Data Fig. 4c. Although this analysis can be performed with the axial-view images, for Fig. 2 we did this through separate experiments with images taken along the y axis, which had the benefits of a reduced optical depth and a smoother intensity profile of the imaging beam. These side-view images are 2D projections of the photofragment position onto the xz plane, and are complicated by the distribution of occupied sites in the optical lattice.

Calculation and parameterization of angular distributions. Supplementary Information details the calculation and parameterization of photodissociation angular distributions used in this work. Supplementary Tables 1–3 list the parameters for all of the theoretical images as well as experimental continuum energies.

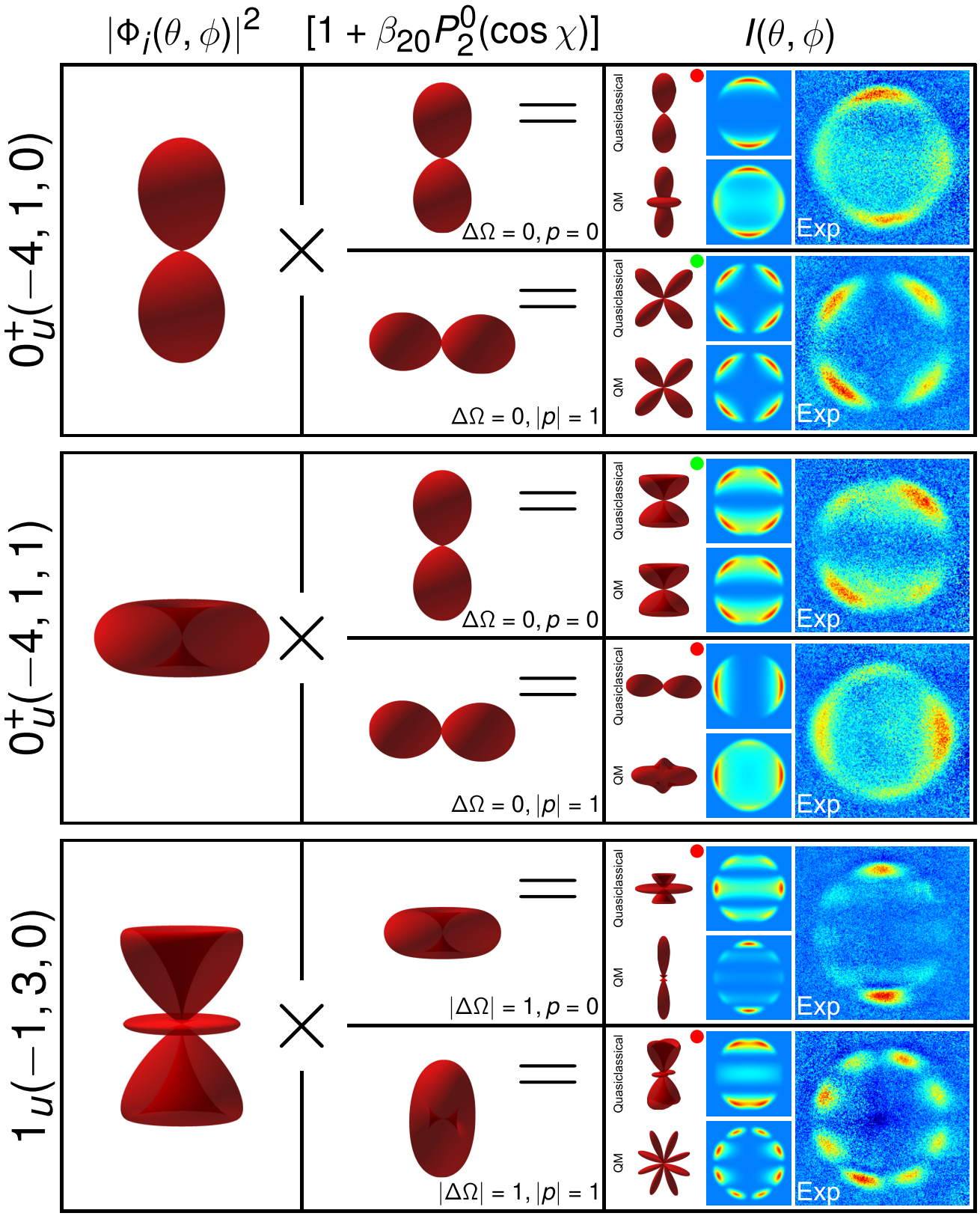
31. Reinaudi, G., Osborn, C. B., McDonald, M., Kotochigova, S. & Zelevinsky, T. Optical production of stable ultracold $^{88}Sr_2$ molecules. *Phys. Rev. Lett.* **109,** 115303 (2012).
32. McGuyer, B. H. *et al.* Nonadiabatic effects in ultracold molecules via anomalous linear and quadratic Zeeman shifts. *Phys. Rev. Lett.* **111,** 243003 (2013).
33. Reinaudi, G., Lahaye, T., Wang, Z. & Guéry-Odelin, D. Strong saturation absorption imaging of dense clouds of ultracold atoms. *Opt. Lett.* **32,** 3143–3145 (2007).
34. Wrede, E., Wouters, E. R., Beckert, M., Dixon, R. N. & Ashfold, M. N. R. Quasiclassical and quantum mechanical modeling of the breakdown of the axial recoil approximation observed in the near threshold photolysis of IBr and Br_2. *J. Chem. Phys.* **116,** 6064–6071 (2002).
35. Apfelbeck, F. *Photodissociation Dynamics of Ultracold Strontium Dimers*. MSc thesis, Ludwig Maximilian University of Munich (2015).
36. O'Keeffe, P. *et al.* A photoelectron velocity map imaging spectrometer for experiments combining synchrotron and laser radiations. *Rev. Sci. Instrum.* **82,** 033109 (2011).

© 2016 Macmillan Publishers Limited. All rights reserved



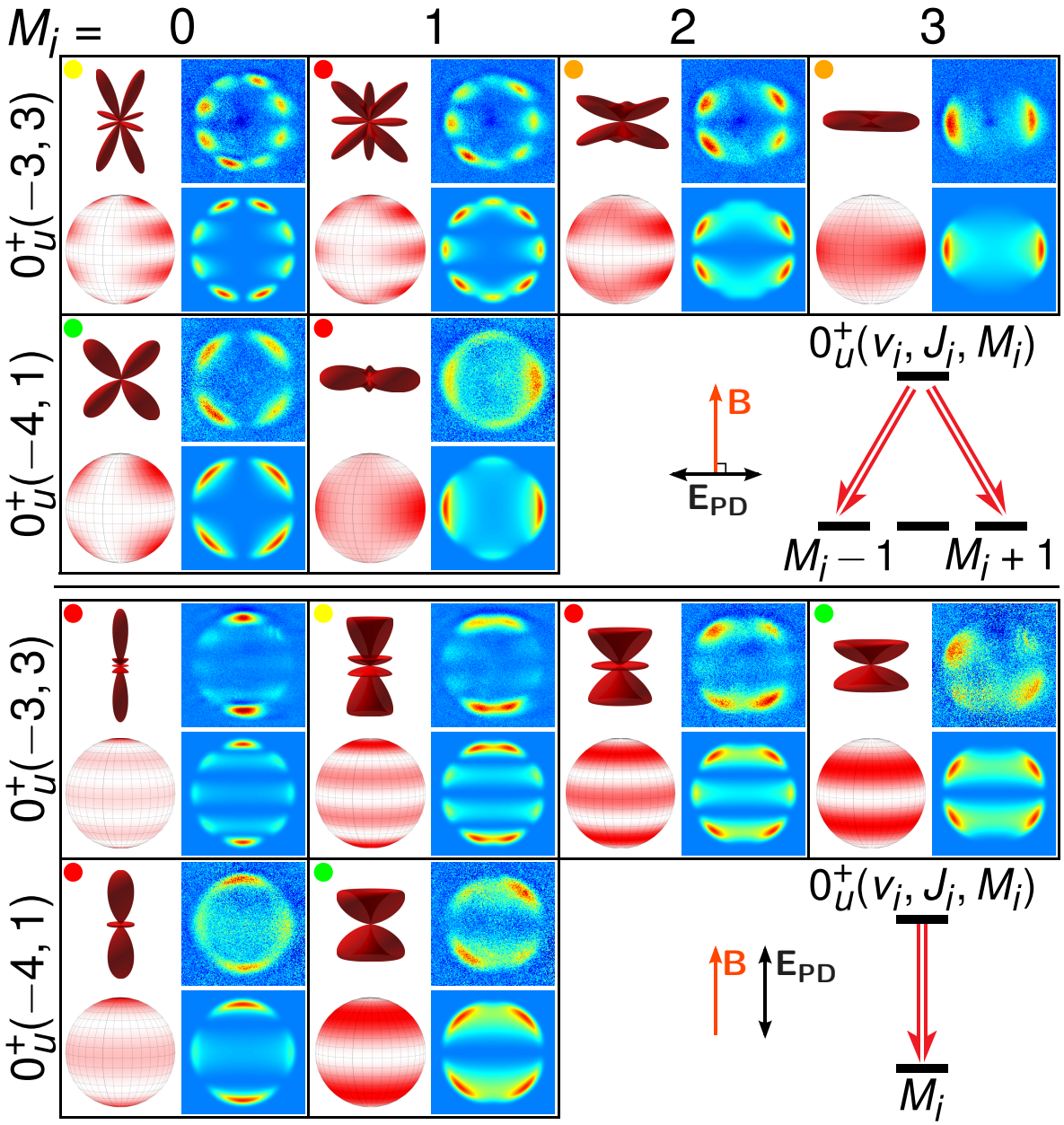
Extended Data Figure 1 | Angular distributions for the M1/E2 photodissociation of $1_g(v_i = -1, J_i = 1, M_i = 0)$ state with $p = 0$ to the ground continuum. Images are arranged as in Fig. 4. The experimental images are labelled by the continuum energy ε/h in MHz. To improve contrast, the strong centre dot from spontaneous decay, as seen in Fig. 3b, was removed before processing and is covered by a box. The theoretical images are calculated using a quantum chemistry model.

© 2016 Macmillan Publishers Limited. All rights reserved



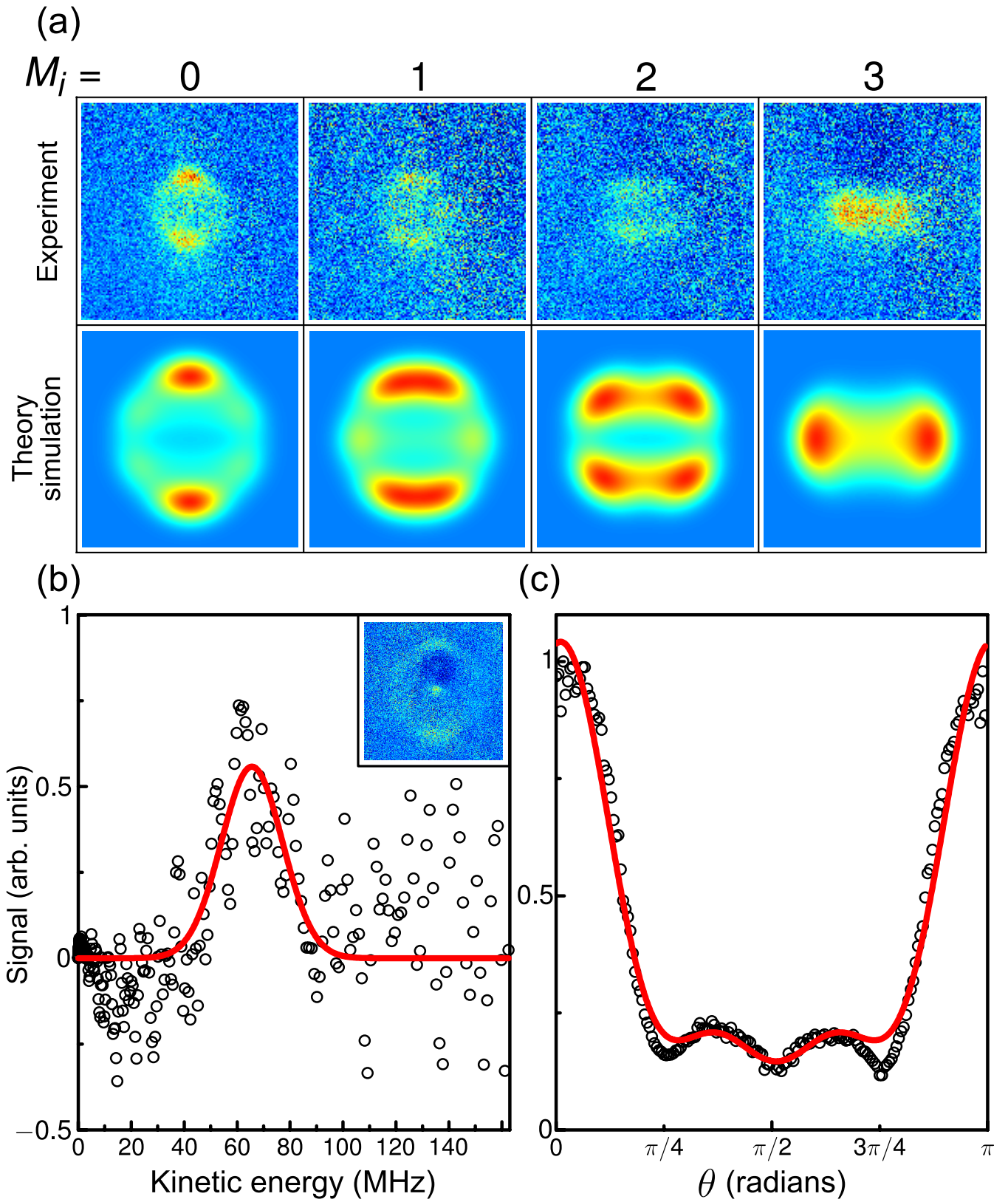
Extended Data Figure 2 | Comparison of quasiclassical and quantum mechanical (QM) theory with experimental (Exp) images for selected cases from Fig. 4 and Extended Data Fig. 3. The quasiclassical predictions follow from equations (3) and (4) assuming $\beta_{20}=2$ for $\Delta\Omega=0$ and -1 for $|\Delta\Omega|=1$. (The quantum mechanical predictions slightly differ from those displayed in Fig. 4 and Extended Data Fig. 3 because they are the full quantum mechanical calculations given in Supplementary Tables 2 and 3.) As before, coloured dots indicate the level of quasiclassical agreement.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 3 | Photodissociation of molecules near the $^1S + {}^3P_1$ threshold to the ground-state continuum. In contrast to Fig. 4, here the initial states are 0_u^+ with $(v_i, J_i) = (-4, 1)$ or $(-3, 3)$ as indicated. These initial states lead to nearly identical distributions as those with the 1_u initial states, contrary to the quasiclassical picture. As before, compatibility with the quasiclassical approximation is indicated by the coloured dots.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Spontaneous photodissociation of molecules prepared in $0_u^+(v_i = -3, J_i = 3, M_i)$ states. a, Absorption images of angular distributions versus M_i. Theoretical simulations using equation (5) are shown underneath. A short expansion time was used to increase visibility. **b**, For quantitative analysis, another image (inset) of the $M_i = 0$ case was taken with a longer expansion time and analysed with the pBasex algorithm. The extracted fragment radial distribution shows a focusing around a certain kinetic energy, which was determined by fitting with a Gaussian (red curve). Correcting for an offset due to the lattice depth[35], this energy corresponds to a shape resonance with a binding energy of -66 ± 3 MHz. **c**, The extracted fragment angular distribution qualitatively matches the calculation (red curve) of equation (5).

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature17974

Single-molecule strong coupling at room temperature in plasmonic nanocavities

Rohit Chikkaraddy[1], Bart de Nijs[1], Felix Benz[1], Steven J. Barrow[2], Oren A. Scherman[2], Edina Rosta[3], Angela Demetriadou[4], Peter Fox[4], Ortwin Hess[4] & Jeremy J. Baumberg[1]

Photon emitters placed in an optical cavity experience an environment that changes how they are coupled to the surrounding light field. In the weak-coupling regime, the extraction of light from the emitter is enhanced. But more profound effects emerge when single-emitter strong coupling occurs: mixed states are produced that are part light, part matter[1,2], forming building blocks for quantum information systems and for ultralow-power switches and lasers[3–6]. Such cavity quantum electrodynamics has until now been the preserve of low temperatures and complicated fabrication methods, compromising its use[5,7,8]. Here, by scaling the cavity volume to less than 40 cubic nanometres and using host–guest chemistry to align one to ten protectively isolated methylene-blue molecules, we reach the strong-coupling regime at room temperature and in ambient conditions. Dispersion curves from more than 50 such plasmonic nanocavities display characteristic light–matter mixing, with Rabi frequencies of 300 millielectronvolts for ten methylene-blue molecules, decreasing to 90 millielectronvolts for single molecules—matching quantitative models. Statistical analysis of vibrational spectroscopy time series and dark-field scattering spectra provides evidence of single-molecule strong coupling. This dressing of molecules with light can modify photochemistry, opening up the exploration of complex natural processes such as photosynthesis[9] and the possibility of manipulating chemical bonds[10].

Creating strongly coupled mixed states from visible light and individual emitters is severely compromised by the hundred-fold difference in their spatial localization. To overcome this, high-quality cavities are used to boost interaction times and enhance coupling strengths. However, in larger cavities the longer round trip for photons to return to the same emitter decreases the coupling, which scales as $g \propto 1/\sqrt{V}$, where V is the effective cavity volume and g is the coupling energy. This coupling has to exceed both the cavity loss rate, κ, and the emitter scattering rate, γ, in order for energy to cycle back and forth between matter and light components, requiring $2g > \gamma, \kappa$ (ref. 11). For cryogenic emitters[5,6] (laser-cooled atoms, vacancies in diamond, or semiconductor quantum dots), the suppressed emitter scattering allows large cavities (with a high quality factor, Q, which is proportional to κ^{-1}) to reach strong coupling. Severe technical challenges, however, restrict the energy, bandwidth, size and complexity of devices. Progress towards room-temperature devices has been limited by the unavoidable increase in emitter scattering, and the difficulty of reducing the volume of dielectric-based microcavities—at wavelength λ and refractive index n—below $V_\lambda = (\lambda/n)^3$. At room temperature, typical scattering rates for embedded dipoles are $\gamma \sim k_B T$, implying that suitable $Q < 100$, which thus requires cavities of less than $10^{-5} V_\lambda$ (Fig. 1a, dark green shaded area).

Improved confinement uses surface plasmons (Fig. 1a), combining oscillations of free electrons in metals with electromagnetic waves[12]. Structured metal films can couple molecular aggregates of high oscillator strength, but far too many molecules are involved for quantum optics. Recent studies have reached 1,000 molecules[13–15]—still far above the one to ten molecules that are needed to access quantum effects at room temperature.

To create such small nanocavities and orient single molecules precisely within them, we use bottom-up nanoassembly. Although field volumes of individual plasmonic nanostructures are too large[1,2], smaller volumes and stronger field enhancements occur within subnanometre gaps between paired plasmonic nanoparticles. We use the promising nanoparticle-on-mirror (NPoM) geometry[16], placing emitters in the gap between nanoparticles and a mirror underneath (Fig. 1b). This gap is accurately controlled to a subnanometre scale using molecular spacers, is easily made by depositing monodisperse metal nanoparticles onto a metal film, and is scalable, repeatable and straightforward to characterize[17,18]. Specifically, we use gold nanoparticles of 40-nm diameter on a 70-nm-thick gold film, separated by a 0.9-nm molecular spacer (see below). The intense interaction between each nanoparticle and its image forms a dimer-like construct with field enhancements of $\sim 10^3$, and an ultralow mode volume. The coupled plasmonic dipolar mode is localized in the gap (Fig. 1b), with the electric field oriented vertically (along the z direction). The resonant wavelength is determined by the nanoparticle size and gap thickness, and can thus be tuned from 600 nm to 1,200 nm (ref. 17).

Several factors are essential in positioning a quantum emitter inside these small gaps. One is to prevent molecular aggregation, which occurs commonly. Another is to ensure that the transition dipole moment, $\boldsymbol{\mu}_m$, is perfectly aligned with the gap plasmon (along the electric field). We use a common dye molecule, methylene blue, with a molecular transition at 665 nm, to which our plasmons are tuned. To avoid aggregation of the dye molecules and to assemble them in the proper orientation, we use the host–guest chemistry of macrocyclic cucurbit[n]uril molecules. These are pumpkin-shaped molecules with varying hollow hydrophobic internal volumes, determined by the number of units in the ring (n), in which guest molecules can sit (Supplementary Fig. 1)[19]. Cucurbit[7]uril is water-soluble and can accommodate only one methylene-blue molecule inside. Encapsulation of methylene blue inside cucurbit[7]uril is confirmed by absorption spectroscopy (Fig. 2a): methylene-blue dimers (shown by the small 'shoulder' peak at 625 nm on the red curve) disappear on mixing low methylene-blue concentrations with cucurbit[7]uril (in a 1:10 molar ratio) (Fig. 2a, blue curve). Control experiments with the smaller cucurbit[5]uril molecules (into which methylene blue cannot fit) do not remove this shoulder peak (Fig. 2a, dashed line), ruling out parasitic binding. Placing single methylene-blue molecules into cucurbit[7]uril thus avoids any aggregation. Carbonyl portals at either end of the 0.9-nm-high cucurbit[n]uril molecules bind them with their rims flat onto the gold surface (Fig. 2b). When a monolayer of cucurbit[7]uril is first deposited on the gold mirror and suitably filled with methylene-blue molecules, gold nanoparticles can bind on top to form the desired filled nanocavity

[1]NanoPhotonics Centre, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK. [2]Melville Laboratory for Polymer Synthesis, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. [3]Department of Chemistry, King's College London, London SE1 1DB, UK. [4]Blackett Laboratory, Department of Physics, Prince Consort Road, Imperial College, London SW7 2AZ, UK.

© 2016 Macmillan Publishers Limited. All rights reserved

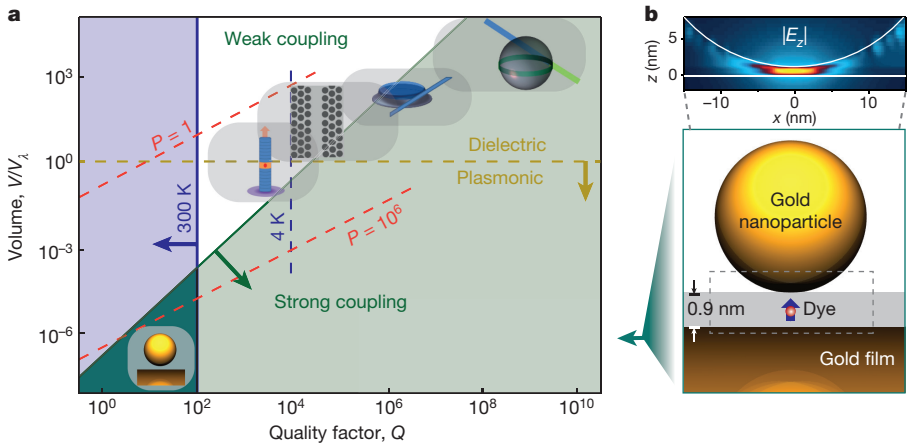


Figure 1 | Comparing single-molecule optical cavities. a, The quality factor, Q, of a nanocavity is plotted against its effective volume, V/V_λ (scaled to $V_\lambda = (\lambda/n)^3$), showing strong-coupling (green arrow), room-temperature (blue arrow), and plasmonic (orange arrow) regimes for single emitters. The icons show realizations of each type of nanocavity: from right, whispering gallery spheres (used as microresonators in filters, sensors and lasers), microdisks, photonic crystals (with possible applications in optical computing), micropillars (used in high-throughput screening), and nanoparticle-on-mirror geometry (NPoM, used here). Purcell factors (P) show emission-rate enhancements. **b**, Diagram of a NPoM. The blue arrow in the gap between the nanoparticle and the mirror locates the transition dipole moment of the emitter. The inset above shows the simulated near-field of the coupled gap plasmon in the dashed box, with maximum electric field enhancement of about 400, oriented vertically (in the z direction).

(Fig. 2b and Supplementary Information), with the methylene-blue molecule aligned vertically in the gap[19]. Previous studies[17] with empty cucurbit[n]urils show that the gap is 0.9 nm, with a refractive index of 1.4.

Dark-field scattering spectra from individual NPoMs show the effect of aligning the emitter in different orientations (Fig. 3a). With $\boldsymbol{\mu}_m$ parallel to the mirror (top; without cucurbit[n]urils the methylene blue lies flat on the metal surface), the resonant scattering plasmonic peak (ω_p) is identical to that of NPoMs without any emitters (ω_0). But with $\boldsymbol{\mu}_m$ perpendicular to the mirror (bottom), the spectra show two split peaks (ω_+ and ω_-) resulting from the strong interaction between emitters and plasmon. We contrast three types of samples. Without dye (Fig. 3b, top), a consistent gap plasmon (ω_p) at 660 ± 10 nm is seen. Small fluctuations in peak wavelength are associated with $\pm$5-nm variations in nanoparticle size (Supplementary Fig. 2). When this NPoM is partially filled with methylene blue inside the cucurbit[7]uril, peaks at 610 nm and 750 nm are seen either side of the absorption peak of methylene blue at ω_0 (Fig. 3b, bottom), corresponding to the formation of hybrid plasmon–exciton ('plexciton') branches, $\omega_\pm = \omega_0 \pm g/2$. This yields a Rabi frequency of $g = 380$ meV, confirmed by full three-dimensional finite-difference time-domain (FDTD) simulations (Supplementary Fig. 3). While some studies[13,14] have shown significant variations in $\omega_\pm$, we obtain highly consistent results, with no spectral wandering observed on individual NPoMs. With dye molecules perpendicular to the plasmon field (without cucurbit[n]urils), only a gap plasmon is seen (Supplementary Fig. 4c). Methylene-blue molecules self-assembling on gold orient flat to the surface, owing to π-stacking interactions between the conjugated phenyl rings and the metal film[20]. Our study thus shows how molecular scaffolding is essential to yield molecular coupling to the gap plasmon.

To map the dispersion curve, we combine scattering spectra from differently sized nanoparticles, plotted according to their detuning from the absorption ('exciton') resonance. Simulations of nanoparticles of 40–60 nm in diameter (Supplementary Fig. 5) show gap plasmons tuning across the exciton. A simple coupled-oscillator model matches the quantum mechanical Jaynes–Cummings picture[13]:

$$\omega_\pm = \frac{1}{2}(\omega_p + \omega_0) \pm \frac{1}{2}\sqrt{g^2 + \delta^2}$$

with plasmon and exciton resonance energies ω_p and ω_0, and detuning energies of $\delta = \omega_p - \omega_0$. Extracting $\omega_\pm$ from the scattering spectra allows ω_p to be calculated (knowing ω_0, which does not show any spectral wandering). This fitting reveals typical anticrossing (mixing) behaviour (Fig. 3c), with $g = 305 \pm 8$ meV at $\delta = 0$. We find $2g/\gamma_{pl} \sim 5$, well into the strong coupling regime. A key figure of merit is the Purcell factor, $P = Q/V$, which characterizes different cavity systems (Fig. 1a). For our plasmonic nanocavities, $P \approx 3.5 \times 10^6$ (Supplementary Fig. 6); this is over an order of magnitude larger than the Purcell factors of state-of-the-art photonic crystal cavities[5], which have reached 10^5, while state-of-the-art planar micropillars[21,22] attain Purcell factors of 3×10^5. The ultralow cavity volume arises here because of the very large field confinement in such nanometre-sized gaps (Supplementary Fig. 9e). Such Purcell factors imply photon emission times below 100 femtoseconds, seen as the $\hbar/g \sim$30-femtosecond Rabi flopping, but very short to measure directly.

To probe single-molecule strong coupling, we systematically decrease the number of methylene-blue molecules by reducing the ratio of methylene blue to cucurbit[7]uril. Previous studies and simple area estimates imply that 100 cucurbit[7]uril molecules lie inside each nanocavity (Supplementary Fig. 9). With the initial 1:10 molar ratio of methylene

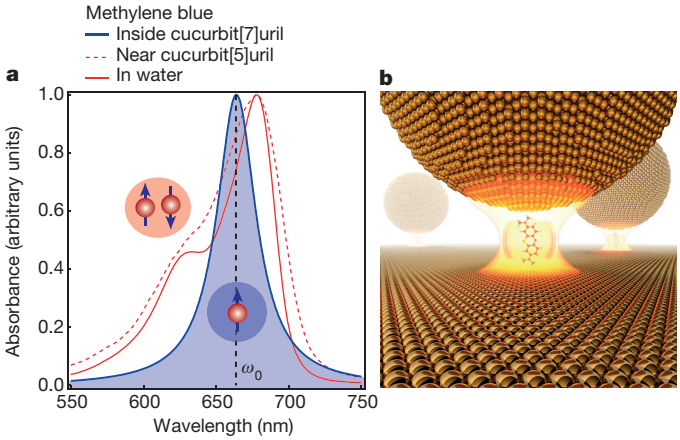


Figure 2 | Plasmonic nanocavity containing a dye molecule. **a**, Absorption spectra of methylene blue in water, with (blue) and without (red) encapsulation in cucurbit[n]urils of different diameters (dashed and solid red lines). Icons show individual molecules (in blue; line centred at ω_0) and paired molecular dimers (in red). **b**, Illustration of a methylene-blue molecule in cucurbit[n]uril, in the nanoparticle-on-mirror geometry used here.

© 2016 Macmillan Publishers Limited. All rights reserved

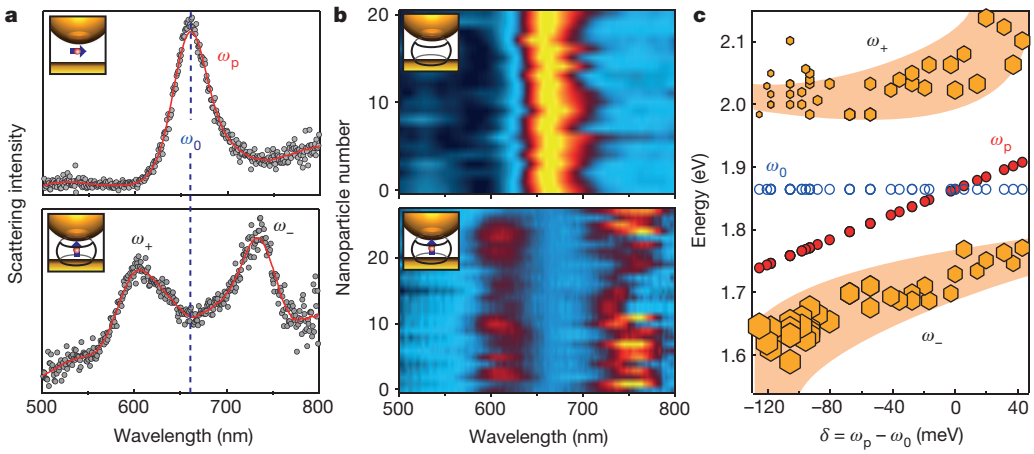


Figure 3 | Strong coupling seen in scattering spectra of individual NPoMs. a, Scattering spectra resulting from isolated NPoMs according to the orientation of the emitter (the methylene-blue dye; see insets). With the dye transition dipole moment, $\boldsymbol{\mu}_m$, oriented parallel to the mirror, the resonant scattering plasmonic peak (ω_p) is identical to that of NPoMs without any emitters. With $\boldsymbol{\mu}_m$ oriented parallel to the mirror, split peaks result from the strong interaction between the emitter and the plasmon. The blue dashed line indicates the dye's absorption wavelength (centred at ω_0). **b**, Comparison of scattering spectra from different NPoMs (see insets), whose gaps are filled by a cucurbit[7]uril monolayer that is empty (top), or encapsulating dye molecules (bottom). **c**, Resonant positions of methylene-blue (ω_0), plasmon (ω_p) and hybrid modes (ω_+ and ω_-) as a function of extracted detuning. The symbol size depicts the amplitude in scattering spectra.

blue:cucurbit[7]uril, the mean number ($\bar{n}$) of methylene-blue molecules within each mode volume is thus 10. We explore many plasmonic nanocavities with a mean dye number of 10 or less (Fig. 4a). From the resulting spectra, we extract coupling strengths at different mean dye numbers, and plot these along with the predicted coupling strength:

$$g_n = \boldsymbol{\mu}_m \sqrt{\frac{4\pi\hbar nc}{\lambda\varepsilon\varepsilon_0 V}}$$

where $\boldsymbol{\mu}_m = 3.8D$ is the transition dipole moment of isolated methylene-blue molecules[23]. The probability of finding each coupling strength (Fig. 4a, colour map) follows the Poisson distribution for n molecules under each nanoparticle. The range of Rabi splittings seen for $\bar{n} = 2.5$ that exceed thermal- and cavity-loss rates at room temperature, is consistent with the idea that our plasmonic nanocavity is supporting single-molecule strong coupling. Reassuringly, the range of Rabi frequencies observed increases as the molecular concentration is reduced, as would be expected

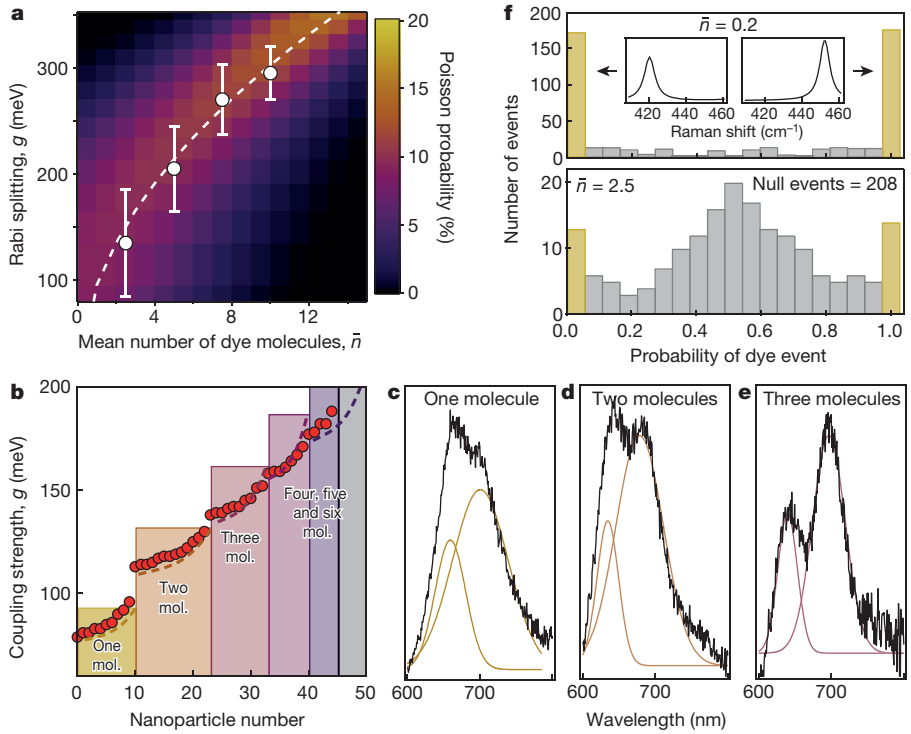


Figure 4 | Rabi splitting from few molecules. a, Energy of Rabi oscillations (g) versus mean number of dye molecules ($\bar{n}$). Experimental (white) points are shown, together with the range of measured coupling strengths (error bars) compared with the theoretical curve (dashed line). The colours represent the Poisson probability distribution of $\bar{n}$. **b**, Coupling strength extracted from different NPoMs in a sample of $\bar{n} = 2.5$. The bars show the theoretical coupling strength obtained from a perfect model; the dashed lines show a random-placement model. **c**–**e**, Scattering spectra for one, two and three molecules (corresponding to **b**), with fits. **f**, Single-molecule probability histograms for $\bar{n} = 0.2$ and 2.5, derived from modified principal-component analysis (Supplementary Fig. 13). The yellow bars show single-molecule events. The insets show the Raman signatures of the two different types of molecular event.

© 2016 Macmillan Publishers Limited. All rights reserved

given that $\Delta g(\bar{n}) \propto \sqrt{\bar{n}+\bar{n}^{1/2}} - \sqrt{\bar{n}-\bar{n}^{1/2}}$ similarly increases, as observed in Fig. 4a (colour map).

Direct proof of single-molecule strong coupling is seen from the coupling strengths extracted from the lowest-density samples ($\bar{n}=2.5$): these show distinct, systematic jumps matching the expected increase in g_n as n rises from one to three dye molecules (Fig. 4b; NPoMs are sorted according to increasing Rabi splitting). The range in each value of g_n arises because single molecules are located at different lateral positions within the gap plasmon, thus coupling with different strengths (predictions are shown as dashed lines in Fig. 4b). Experimentally, we find excellent agreement (with no fitting parameters), showing that a single methylene-blue molecule in our nanocavities gives Rabi splittings of 80–95 meV. Further, we plot the scattering spectrum from $n=1$–3 molecules, revealing clear increases in coupling strength. Additional proof of the single-molecule strong coupling is seen from the anticrossing of plasmon and exciton modes for the subset with $n=1$ (Supplementary Fig. 20).

For weakly coupled single molecules, emitted fluorescence should follow the Purcell factor[24–26]. However, such measurements fail here in the strong-coupling regime, because resonantly pumping the molecular absorption also generates strong surface-enhanced resonant Raman scattering (SERRS)—consisting of sharp lines with a strong background—that cannot be uniquely separated from photoluminescence (Supplementary Fig. 11). This also obscures the $g^{(2)}$ measurements that are typically used to confirm single-photon emission from individual chromophores. Here we find extremely strong emission—even though the dye molecules are within 0.5 nm of absorptive gold[27]—owing to the high radiative efficiency of our nanocavities. We harvest these strong SERRS peaks to construct 'chemical' $g^{(2)}_{\mathrm{ch}}$ values, by using the well established bianalyte technique with a second near-identical but distinguishable molecule to prove single-molecule statistics (Fig. 4f and Supplementary Figs 15–17). As clearly evident, at the lowest concentrations two molecules are almost never found at the same time, and we are truly in the single-molecule regime. Although this does not guarantee direct correlation with single-molecule strong-coupling situations, it does prove the statistical probability of single molecules at this concentration. Convincing proof of the presence of single molecules is also provided by the spectral diffusion of vibrational lines in time-series SERRS scans from nanoparticles exhibiting single-molecule strong coupling (Supplementary Figs 18 and 19).

We have succeeded in combining the gap plasmon with oriented host–guest chemistry in aqueous solution to create enormous numbers of strongly coupled, few-molecule nanocavities at room temperature, in ambient conditions, and which are optically addressable. We envisage numerous applications, including single-photon emitters, photon blockades[9], quantum chemistry[28–30], nonlinear optics, and tracked or directed molecular reactions.

Received 16 July 2015; accepted 1 April 2016.
Published online 13 June 2016.

1. Tame, M. S. *et al.* Quantum plasmonics. *Nature Phys.* **9,** 329–340 (2013).
2. Koenderink, A. F., Alù, A. & Polman, A. Nanophotonics: shrinking light-based technology. *Science* **348,** 516–521 (2015).
3. Sato, Y. *et al.* Strong coupling between distant photonic nanocavities and its dynamic control. *Nature Photon.* **6,** 56–61 (2012).
4. Liu, X. *et al.* Strong light–matter coupling in two-dimensional atomic crystals. *Nature Photon.* **9,** 30–34 (2015).
5. Yoshie, T. *et al.* Vacuum Rabi splitting with a single quantum dot in a photonic crystal nanocavity. *Nature* **432,** 200–203 (2004).
6. Thompson, J. D. *et al.* Coupling a single trapped atom to a nanoscale optical cavity. *Science* **340,** 1202–1205 (2013).
7. Faraon, A. *et al.* Coherent generation of non-classical light on a chip via photon-induced tunnelling and blockade. *Nature Phys.* **4,** 859–863 (2008).
8. Gröblacher, S. *et al.* An experimental test of non-local realism. *Nature* **446,** 871–875 (2007).
9. Coles, D. M. *et al.* Strong coupling between chlorosomes of photosynthetic bacteria and a confined optical cavity mode. *Nature Commun.* **5,** 5561 (2014).
10. Shalabney, A. *et al.* Coherent coupling of molecular resonators with a microcavity mode. *Nature Commun.* **6,** 5981 (2015).
11. Törmä, P. & Barnes, W. L. Strong coupling between surface plasmon polaritons and emitters: a review. *Rep. Prog. Phys.* **78,** 013901 (2015).
12. Novotny, L. & Hecht, B. *Principles of Nano-Optics* (Cambridge Univ. Press, 2006).
13. Zengin, G. *et al.* Realizing strong light-matter interactions between single-nanoparticle plasmons and molecular excitons at ambient conditions. *Phys. Rev. Lett.* **114,** 157401 (2015).
14. Zengin, G. *et al.* Approaching the strong coupling limit in single plasmonic nanorods interacting with J-aggregates. *Sci. Rep.* **3,** 3074 (2013).
15. Schlather, A. E., Large, N., Urban, A. S., Nordlander, P. & Halas, N. J. Near-field mediated plexcitonic coupling and giant Rabi splitting in individual metallic dimers. *Nano Lett.* **13,** 3281–3286 (2013).
16. Ciracì, C. *et al.* Probing the ultimate limits of plasmonic enhancement. *Science* **337,** 1072–1074 (2012).
17. de Nijs, B. *et al.* Unfolding the contents of sub-nm plasmonic gaps using normalising plasmon resonance spectroscopy. *Faraday Discuss.* **178,** 185–193 (2015).
18. Benz, F. *et al.* Nanooptics of molecular-shunted plasmonic nanojunctions. *Nano Lett.* **15,** 669–674 (2015).
19. Kasera, S., Herrmann, L. O., del Barrio, J., Baumberg, J. J. & Scherman, O. A. Quantitative multiplexing with nano-self-assemblies in SERS. *Sci. Rep.* **4,** 6785 (2014).
20. Netzer, F. P. & Ramsey, M. G. Structure and orientation of organic molecules on metal surfaces. *Crit. Rev. Solid State Mater. Sci.* **17,** 397–475 (1992).
21. Vahala, K. J. Optical microcavities. *Nature* **424,** 839–846 (2003).
22. Khitrova, G., Gibbs, H. M., Kira, M., Koch, S. W. & Scherer, A. Vacuum Rabi splitting in semiconductors. *Nature Phys.* **2,** 81–90 (2006).
23. Patil, K., Pawar, R. & Talap, P. Self-aggregation of methylene blue in aqueous medium and aqueous solutions of Bu4NBr and urea. *Phys. Chem. Chem. Phys.* **2,** 4313–4317 (2000).
24. Akselrod, G. M. *et al.* Probing the mechanisms of large Purcell enhancement in plasmonic nanoantennas. *Nature Photon.* **8,** 835–840 (2014).
25. Anger, P., Bharadwaj, P. & Novotny, L. Enhancement and quenching of single-molecule fluorescence. *Phys. Rev. Lett.* **96,** 113002 (2006).
26. Kinkhabwala, A. *et al.* Large single-molecule fluorescence enhancements produced by a bowtie nanoantenna. *Nature Photon.* **3,** 654–657 (2009).
27. Kravtsov, V., Berweger, S., Atkin, J. M. & Raschke, M. B. Control of plasmon emission and dynamics at the transition from classical to quantum coupling. *Nano Lett.* **14,** 5270–5275 (2014).
28. Hutchison, J. A., Schwartz, T., Genet, C., Devaux, E. & Ebbesen, T. W. Modifying chemical landscapes by coupling to vacuum fields. *Angew. Chem. Int. Ed.* **51,** 1592–1596 (2012).
29. Galego, J., Garcia-Vidal, F. J. & Feist, J. Cavity-induced modifications of molecular structure in the strong coupling regime. *Phys. Rev. X* **5,** 041022 (2015)
30. Feist, J. & Garcia-Vidal, F. J. Extraordinary exciton conductance induced by strong coupling. *Phys. Rev. Lett.* **114,** 196402 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge financial support from the UK's Engineering and Physical Sciences Research Council (grants EP/G060649/1, EP/N020669/1, EP/L027151/1 and EP/l012060/1) and the European Research Council (grant LINASS 320503). This study was partially supported by the Air Force Office of Scientific Research (AFOSR); the European Office of Aerospace Research and Development (EOARD) is also acknowledged. R.C. acknowledges support from the Dr. Manmohan Singh scholarship from St John's College, University of Cambridge. F.B. acknowledges support from the Winton Programme for the Physics of Sustainability. S.J.B. acknowledges support from the European Commission for a Marie Curie Fellowship (NANOSPHERE, 658360).

Author Contributions J.J.B. and R.C. conceived and designed the experiments. R.C. performed the experiments with input from F.B. and B.d.N. R.C. and A.D. carried out the simulation and the analytical modelling with input from J.J.B., P.F., O.H. and E.R. R.C. and J.J.B. analysed the data. S.J.B. and O.A.S. synthesized cucurbit[*n*]uril and provided input on the fabrication and characterization of samples. R.C. and J.J.B. wrote the manuscript with input from all authors.

Author Information Data supporting this paper are available at https://www.repository.cam.ac.uk/handle/1810/254579. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.J.B. (jjb12@cam.ac.uk).

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18284

Lanthanum–catalysed synthesis of microporous 3D graphene–like carbons in a zeolite template

Kyoungsoo Kim[1], Taekyoung Lee[1,2], Yonghyun Kwon[1,2], Yongbeom Seo[1], Jongchan Song[3], Jung Ki Park[3], Hyunsoo Lee[1], Jeong Young Park[1,4], Hyotcherl Ihee[1,2], Sung June Cho[5] & Ryong Ryoo[1,2]

Three-dimensional graphene architectures with periodic nanopores—reminiscent of zeolite frameworks—are of topical interest because of the possibility of combining the characteristics of graphene with a three-dimensional porous structure[1–6]. Lately, the synthesis of such carbons has been approached by using zeolites as templates and small hydrocarbon molecules that can enter the narrow pore apertures[7–15]. However, pyrolytic carbonization of the hydrocarbons (a necessary step in generating pure carbon) requires high temperatures and results in non-selective carbon deposition outside the pores. Here, we demonstrate that lanthanum ions embedded in zeolite pores can lower the temperature required for the carbonization of ethylene or acetylene. In this way, a graphene-like carbon structure can be selectively formed inside the zeolite template, without carbon being deposited at the external surfaces. X-ray diffraction data from zeolite single crystals after carbonization indicate that electron densities corresponding to carbon atoms are generated along the walls of the zeolite pores. After the zeolite template is removed, the carbon framework exhibits an electrical conductivity that is two orders of magnitude higher than that of amorphous mesoporous carbon. Lanthanum catalysis allows a carbon framework to form in zeolite pores with diameters of less than 1 nanometre; as such, microporous carbon nanostructures can be reproduced with various topologies corresponding to different zeolite pore sizes and shapes. We demonstrate carbon synthesis for large-pore zeolites (FAU, EMT and beta), a one-dimensional medium-pore zeolite (LTL), and even small-pore zeolites (MFI and LTA). The catalytic effect is a common feature of lanthanum, yttrium and calcium, which are all carbide-forming metal elements. We also show that the synthesis can be readily scaled up, which will be important for practical applications such as the production of lithium-ion batteries and zeolite-like catalyst supports.

Zeolites are a family of microporous crystalline aluminosilicate materials, which fall into more than 200 structural types[16]. Each structural type is distinguished by its unique pore structure—for example, in terms of its pore diameters, shapes and connectivity[17]. The pore diameters are typically between 0.3 nm and 1.3 nm. Another important characteristic of zeolites is their ion-exchange capacity[18,19]. Zeolite frameworks contain cations to compensate for the negative charge at the aluminiums in the tetrahedral silicate framework. The cations—which, as synthesized, are normally sodium or ammonium ions—can be exchanged with other cations through a solution-based conventional ion-exchange process.

In recent years, zeolites have attracted attention as a template for carbon synthesis[7–15,20,21]. The pores in many zeolites have diameters appropriate to accommodating fullerene and carbon nanotubes, and are interconnected along the smoothly curved surface to form a three-dimensional (3D) network that is open to the exterior. In principle, such a nanoporous system should be ideal as a template for synthesizing a 3D graphene architecture[8–11]. However, the zeolite pores are too small to accommodate bulky molecular compounds, such as sucrose, polyaromatic compounds, and furfuryl alcohol, which are commonly used for carbon synthesis with mesoporous silica templates[20,21]. Small molecules, such as ethylene and acetylene, are desirable as a carbon source for achieving successful carbonization within the zeolite pores. But carbonization of these small hydrocarbons generally requires high-temperature reactions to fix the carbon source inside the pores. At such high temperatures, the reactions tend to occur non-selectively on the external surfaces as well as on the internal pore walls[13–15]. This often results in coke being deposited at the external surfaces, causing serious diffusion limitations into the pores.

Here we tackled this problem by using La^{3+} ions. We intuited that such a transition-metal element would bond with olefins, acetylenes and aromatic compounds through a d–π coordination. If so, then the d–π interactions should stabilize ethylene and the pyrocondensation intermediates to form a carbon framework in zeolite. Then, we would expect carbonization to occur selectively inside the La^{3+}-containing zeolite pores.

To test this hypothesis, we carried out ion exchange of an Na^{+}-containing form of the zeolite faujasite-Y (FAU-Y; that is, NaY zeolite) with La^{3+}. We heated the resulting LaY zeolite under carbon-synthesis conditions using ethylene gas for 1 hour at different temperatures (see Methods). We analysed the amount of carbon deposition at each temperature by thermogravimetry, and plotted the analysis data as a function of temperature (Extended Data Fig. 1). We also compared these LaY data with the results obtained from other cation-containing forms of the zeolite, such as NaY and HY. The data indicate that the LaY, NaY and HY zeolite samples all show rapid carbon deposition at 800 °C. However, as the temperature decreases, the different ionic forms behave dramatically differently: at 600 °C, the LaY zeolite is still active as a carbon-deposition template, whereas both NaY and HY lose this function almost completely. This result highlights a catalytic effect of lanthanum on carbonization. Usually, in carbon synthesis, the proton form of zeolite is preferred as a template. This is due to the presence of Lewis and Brønsted acid sites that can catalyse the pyrocondensation of hydrocarbons into polymeric coke species[22]. But carbon deposition in LaY occurs more than 20 times faster than in such an acidic HY zeolite (based on our chosen ethylene flow for 1 hour at 600 °C). The ethylene flow can also be safely prolonged until all internal pores are fully saturated with carbon; the deposition of any amorphous or graphitic carbon on external surfaces is still prevented (Extended Data Fig. 2).

We investigated the carbon structure using solid-state, magic-angle spinning ^{13}C nuclear magnetic resonance (NMR) spectroscopy after the deposition of ^{13}C-labelled carbon in the LaY zeolite

[1]Center for Nanomaterials and Chemical Reactions, Institute for Basic Science (IBS), Daejeon 305-701, Korea. [2]Department of Chemistry, KAIST, Daejeon 34141, Korea. [3]Department of Chemical and Biomolecular Engineering, KAIST, Daejeon 34141, Korea. [4]Graduate School of EEWS, KAIST, Daejeon 34141, Korea. [5]Clean Energy Technology Laboratory and Department of Chemical Engineering, Chonnam National University, Gwangju 61186, Korea.

© 2016 Macmillan Publishers Limited. All rights reserved

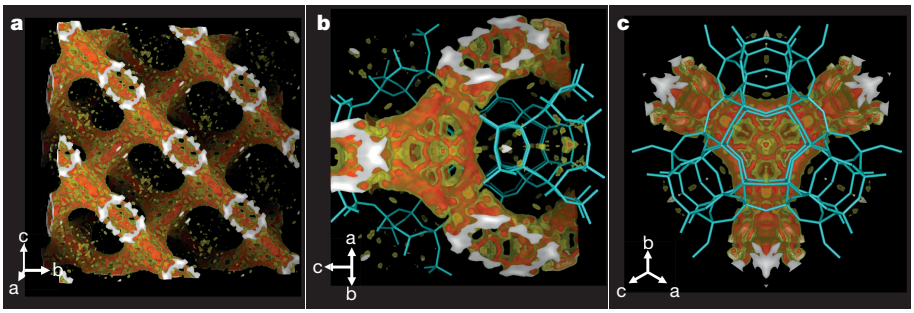


Figure 1 | Electron-density map of the supercage of zeolite FAU after carbon deposition. a, Three-dimensional electron-density map of the carbon framework formed at 600 °C, excluding the zeolite framework. **b**, **c**, Enlarged images of the electron-density map, including the zeolite framework (cyan), from different viewpoints: along the (110) axis (**b**) and the (111) axis (**c**). The electron-density map corresponds to the electron-density difference between zeolite and the carbon–zeolite composite, and was obtained using the difference Fourier method. The iso-surface level of the electron density is set to 0.25 electrons per Å^3 (yellow) and 0.35 electrons per Å^3 (red). White areas represent cross-sectional cuts.

(Extended Data Fig. 3). The NMR spectrum exhibits two slightly separated peaks (at 123 and 129 parts per million, p.p.m.). These NMR peaks can be interpreted as sp^2 carbon species in six-membered, and in five- or seven-membered, carbon rings respectively[23]. Thus, all carbon atoms in the zeolite–carbon composite sample have an sp^2 hybridized bonding nature within the detection limit of the ^{13}C NMR spectroscopy.

Here, the question is whether the carbon structure is built systematically like a 3D graphene along smoothly curved surfaces on pore walls, or exists randomly in the template pore volume. We sought the answer by studying X-ray diffraction (XRD) data of large single crystals of FAU after carbon deposition. Figure 1a shows an electron-density map of atoms that were brought into the zeolite micropore (designated the 'supercage') during carbon deposition. We obtained this map by using the difference Fourier method with X-ray single-crystal diffraction data, which we collected after fully dehydrating the zeolite–carbon composite sample (to exclude moisture in the supercage) by flowing nitrogen gas at 600 °C and then placing the sample in a vacuum at 350 °C (see Methods). All of the electron densities in the supercage can thus be attributed to carbonaceous frameworks.

The electron densities indicate diffused atomic positions in the wide section of the supercage; these atomic positions correspond to a hexagonal ring of carbon atoms, as in a graphene net (Fig. 1b, c). In particular, the density map at cross-sectional cuts exhibits hollow images, indicating that the carbon atoms are systematically deposited along the zeolite supercage surface. However, in the narrow space between adjoining supercages, the electron densities are more diffuse and crowded. Some of the density portions are too close to assign carbon–carbon bonding. This can be interpreted as the average electron-density map superimposing various atomic positions over many identical pore necks—in other words, there is high static disorder. Because of the severe disorder and fractional occupancy, the exact single-crystal structure was difficult to solve unless constraints were used in the refinement process (Extended Data Table 1, Extended Data Fig. 4, and Methods).

The carbon framework obtained at 600 °C can be separated from the template, by using a mixture of hydrogen fluoride and hydrochloric acid to remove the zeolite. The carbon thus recovered exhibits a narrow distribution of pore diameters in the micropore region, corresponding to the thickness of the template walls. The carbon, however, exhibits only poorly resolved XRD peaks and transmission electron microscope (TEM) lattice fringes, indicating that the pores are not well ordered. The loss of pore order seems to result from insufficient formation of carbon–carbon bonds in the narrow necks between adjacent supercages at 600 °C. To obtain a carbon product with highly ordered pores, the carbon–zeolite composite needs to be heated to 850 °C after the carbon-deposition step at 600 °C. This heat treatment involves a small lattice contraction of the zeolite and loss of more than half of the XRD peaks (Supplementary Fig. 1), indicating that the zeolite template behaves like a shrinking mould to allow a rigid carbon framework to form.

The final carbon product—which is liberated from LaY after heating at 850 °C—is an exact replica of the zeolite pore structure. This carbon exhibits highly (for a microporous carbon) well ordered structures in TEM images and powder XRD patterns (Fig. 2 and Supplementary Fig. 2). The TEM images show no carbon deposition on external surfaces. Approximately 80% of the zeolite pores are replicated with carbon (see Supplementary Discussion and Supplementary Fig. 3 for a more detailed quantitative analysis of the 'quality' of carbon replication). This carbon exhibits high thermal stability in air, as compared with mesoporous carbons that are composed of amorphous frameworks[24], or with other zeolite-templated carbons that are synthesized by two-step carbon infiltration into H^+ or Na^+ zeolites (Extended Data Fig. 5). The thermal stability of the carbon is comparable to that of graphene nanosheets. The high thermal stability and well ordered structure can be attributed to nothing but the effect of lanthanum on carbon deposition. To check this effect, we carried out La^{3+}-ion-exchange into EMT and beta zeolites, and used these zeolites as templates. The resulting carbon frameworks also exhibit high thermal stability and a highly ordered microporous structure (Fig. 2, Supplementary Fig. 2 and Extended Data Fig. 5).

We investigated the possibility of observing a graphene-like atomic arrangement by means of a high-resolution TEM instrument (see Methods), but we failed to obtain direct atomic images. The carbon framework was instantly damaged under the atomic-scale observation condition (which requires an electron beam of very high intensity), even when the electron-acceleration voltage was reduced to 80 kV. In an alternative attempt, we took a selected-area electron-diffraction (SAED) pattern of the carbon synthesized using LaY (Fig. 2e). The SAED pattern showed two low-intensity diffraction rings at the same Bragg angles as graphene (100) and (110) reflections[25], unlike the SAED pattern of amorphous carbon. We interpret the SAED pattern of the LaY-templated carbon as revealing random orientations of six-membered carbon rings existing in a curved, single-layer graphene-like structure. This sp^2 carbon character is confirmed by the NMR spectrum (Extended Data Fig. 3). Moreover, a Raman spectrum shows a strong G-band in addition to a D-band (Extended Data Fig. 6), much like in previous reports on curved nanographene samples[9,26]. The G-band is upshifted from the position of graphite; such a shift is often attributed to a curvature in the graphene structure[9,26]. Another notable feature of the LaY-templated carbon is high electrical conductivity (Fig. 2f and Extended Data Fig. 7). We investigated local electrical conductance using conductive probe atomic force microscopy. The result indicates that electrical conductivity of the LaY-templated carbon is two orders of magnitudes higher than that of the mesoporous carbon CMK-3, which has an amorphous framework[24].

Given these results, we tested the possibility of using other metal ions for ion exchange. We chose Y^{3+} and Ca^{2+}, because these metal

© 2016 Macmillan Publishers Limited. All rights reserved

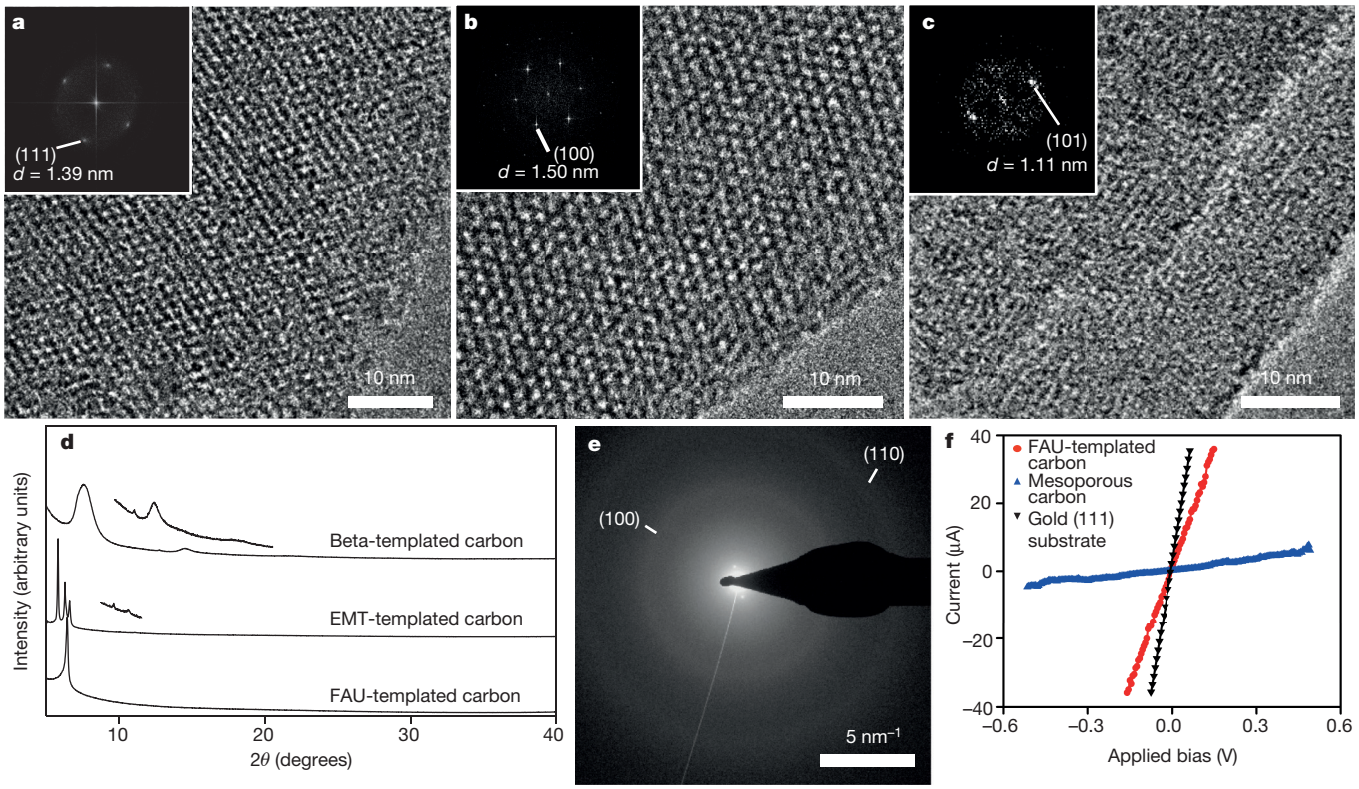


Figure 2 | Structures of 3D graphene-like microporous carbons. **a–c**, TEM images (main pictures) and Fourier diffractograms (insets) of template-free carbon, generated using a template of La^{3+}-ion-exchanged FAU (**a**), EMT (**b**) or beta (**c**) zeolites. The images reveal an ordered pore structure without any external carbon. **d**, Powder XRD patterns from the carbon samples, measured using synchrotron radiation. The patterns reveal the highly ordered microporous structure of the carbons, corresponding to the pore structure of the template. **e**, Electron-diffraction pattern of a selected area from the FAU-templated carbon, showing low-intensity rings corresponding to graphene (100) and (110) reflections. **f**, Current–voltage curves for FAU-templated carbon and CMK-3 mesoporous carbon on a gold (111) substrate, measured by conductive probe atomic force microscopy.

ions are known to interact via coordination bonding with the carbon framework, using electron-donation/back-donation mechanisms[27,28] in a similar way to La^{3+}. Indeed, the exchange of Y^{3+} or Ca^{2+} with ions in FAU-Y dramatically increases the rate of carbon deposition at 600 °C, much as does the LaY zeolite (Supplementary Fig. 4). The final carbon products from these zeolite templates exhibit highly ordered microporous structures. Another critical factor affecting carbon deposition is that water vapour should be fed into the ethylene gas stream. Without water vapour, none of the La^{3+}-, Y^{3+}- and Ca^{2+}-ion-exchanged zeolites shows sufficient carbon deposition. We speculate that this phenomenon could be related to the production of carbides, which the three metal elements used here can form. Typical carbide formation in the bulk state requires high-temperature treatments in an electric arc furnace. However, when metal ions are atomically dispersed in a zeolite framework, a carbide might form even at 600 °C. If so, then as the carbide reacts with water vapour, active carbonaceous species might be generated to construct carbon frameworks.

Acetylene gas can be used instead of ethylene to construct carbon frameworks on the ion-exchanged zeolites. Because acetylene is more reactive than ethylene, carbon deposition can be accomplished at temperatures as low as 340 °C (Supplementary Fig. 5). In addition, the smaller molecular size of acetylene enables uniform infiltration of carbon—even in zeolites with one-dimensional channels (for example, LTL zeolite) or small pore mouths (LTA and MFI zeolites), which have been difficult to use as carbon templates[12,13] (Fig. 3). The LTL zeolite has a one-dimensional (1D) undulating channel, with narrow sections of diameter 0.71 nm and wide sections of diameter 1.24 nm, which alternate with a 0.48-nm periodicity. Accordingly, undulating carbon tubes can be synthesized inside the La^{3+}-exchanged LTL zeolite. When the template walls are removed after heating at 850 °C, the carbon tubes self-assemble to form a bundle (Fig. 3a, b). In contrast, if we use the Na^{+} or H^{+} form of LTL zeolite, carbon deposition occurs only at the external surfaces of the template (Extended Data Fig. 8).

Meanwhile, in the Ca^{2+}-exchanged LTA zeolite, the pore diameter is 1.14 nm; the pores are interconnected to a 3D network, but the pore mouths (of diameter 0.5 nm) are too narrow to have previously considered using LTA as a carbon template. Nonetheless, our results show that carbon infiltrates quite uniformly throughout the entire volume of this zeolite. The final carbon product, liberated from the template, exhibits the crystal morphology of the zeolite template (Fig. 3c), and lattice fringes (Fig. 3e). The carbon crystal can be easily crushed by hand rubbing. The crushed crystal surfaces indicate that the entire volume of the zeolite crystal is used for carbon synthesis (Fig. 3d). Notably, the carbons obtained from the LTL and LTA zeolites can be dispersed in *N*-methylpyrrolidone (NMP); the solutions show photoluminescence, indicating that they are soluble in organic solvents (Fig. 3f). Moreover, the LTA-templated carbon recrystallizes when isopropyl alcohol is added to the NMP solution, indicating that the carbon products show van der Waals packing of carbon nanotubes or carbon dots.

Compared with the LTL and LTA zeolites, the MFI zeolite is somewhat more difficult to use to accomplish carbon synthesis. This is because this zeolite has only narrow channels (<0.56 nm in diameter), without bulged sections. The channels are too narrow to accommodate even C_{60} fullerene. Nevertheless, our results show that these narrow pores can still be used as a template for a carbon nanostructure, and that the morphology of the resulting carbons closely resembles that of the template. The carbon exhibited a sharp peak centred at 0.49 nm in the pore size distribution (Supplementary Fig. 6). This corresponds to

© 2016 Macmillan Publishers Limited. All rights reserved

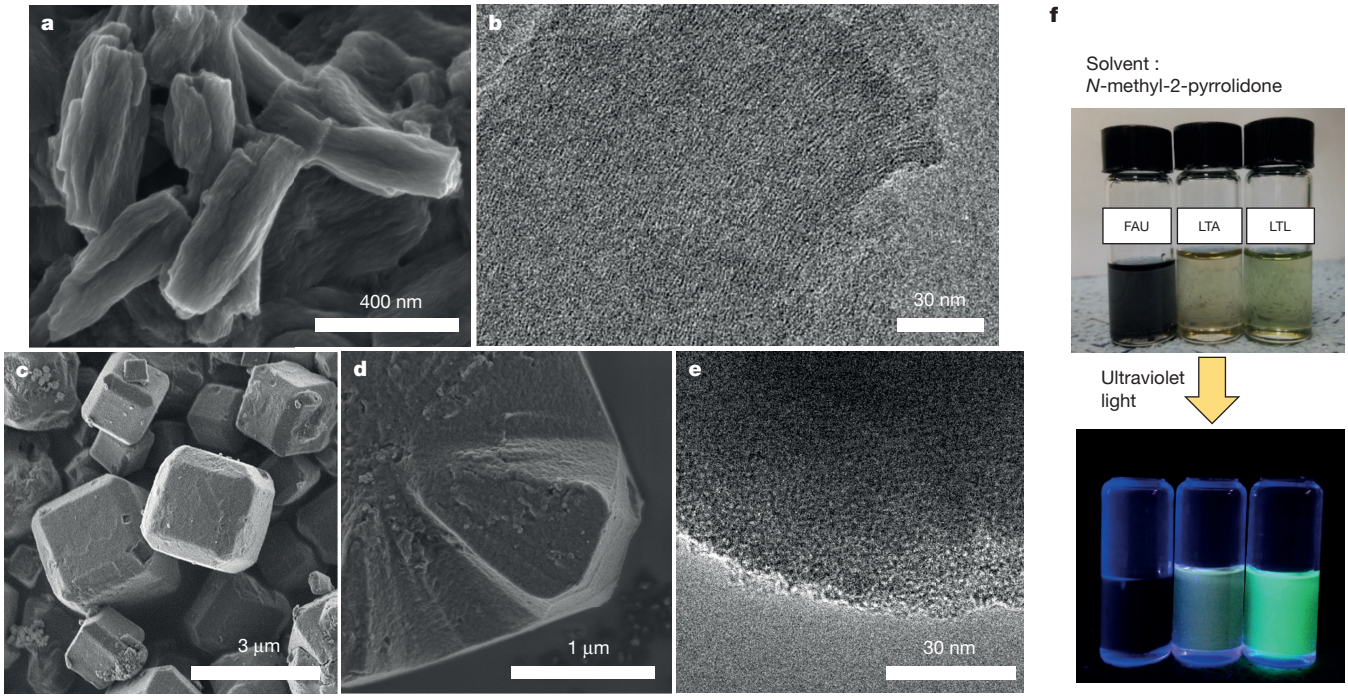


Figure 3 | Carbon from a 1D-channel LTL zeolite, and from a small-pore LTA zeolite. a, Scanning electron microscope (SEM) and **b**, TEM images of LTL-templated carbon, revealing morphologies corresponding to a bundle of 1D channels. **c**, **d**, SEM and **e**, TEM images of LTA-templated carbon, exhibiting zeolite-like crystal morphologies and pore order; these soft carbon crystals can be easily crushed. The SEM image of the broken crystal surfaces (**d**) shows that carbon synthesis uses the entire volume of the zeolite crystal. **f**, Photographs of the LTA- and LTL-templated carbons dispersed in NMP solution, in comparison with FAU. The carbon products from the LTA and LTL templates are soluble and exhibit photoluminescence under ultraviolet light, but the carbon from the FAU template is insoluble. In the latter case, the carbon is synthesized in the form of a 3D porous network that extends over several supercages, rendering it insoluble in organic solvents. However, in the case of the LTA and LTL zeolites, the carbon is obtained as quantum dots or nanotubes; these small carbon objects, which are like fullerenes and carbon nanotubes, are soluble in organic solvents.

the thickness of the MFI pore walls, indicating the formation of rigid carbon nanostructures inside the narrow zeolite channel.

Making graphene with 3D periodic nanoporous architectures promises a range of useful applications, such as in batteries and catalysts[29,30], but has not yet seen full success owing to the lack of efficient synthetic strategies. Our protocol, with its pore-selective carbon filling at decreased temperatures, can be readily scaled up for studies requiring bulk quantities of carbon (Extended Data Fig. 9). Moreover, the high electrical conductivity of the resulting carbon frameworks will be useful in battery applications (Supplementary Fig. 7).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 October 2015; accepted 20 April 2016.
Published online 29 June 2016.

1. Han, S., Wu, D., Li, S., Zhang, F. & Feng, X. Porous graphene materials for advanced electrochemical energy storage and conversion devices. *Adv. Funct. Mater.* **26,** 849–864 (2014).
2. Jiang, L. & Fan, Z. Design of advanced porous graphene materials: from graphene nanomesh to 3D architectures. *Nanoscale* **6,** 1922–1945 (2014).
3. Cao, X., Yin, Z. & Zhang, H. Three-dimensional graphene materials: preparation, structures and application in supercapacitors. *Energy Environ. Sci.* **7,** 1850–1865 (2014).
4. Vanderbilt, D. & Tersoff, J. Negative-curvature fullerene analog of C_{60}. *Phys. Rev. Lett.* **68,** 511–513 (1992).
5. Mackay, A. L. & Terrones, H. Diamond from graphite. *Nature* **352,** 762 (1991).
6. Lenosky, T., Gonze, X., Teter, M. & Elser, V. Energetics of negatively curved graphitic carbon. *Nature* **355,** 333–335 (1992).
7. Ma, Z., Kyotani, T. & Tomita, A. Preparation of a high surface area microporous carbon having the structural regularity of Y zeolite. *Chem. Commun.* **23,** 2365–2366 (2000).
8. Ma, Z., Kyotani, T., Liu, Z., Terasaki, O. & Tomita, A. Very high surface area microporous carbon with a three-dimensional nano-array structure: synthesis and its molecular structure. *Chem. Mater.* **13,** 4413–4415 (2001).
9. Nishihara, H. *et al.* A possible buckybowl-like structure of zeolite templated carbon. *Carbon* **47,** 1220–1230 (2009).
10. Nueangnoraj, K. *et al.* Formation of crosslinked-fullerene-like framework as negative replica of zeolite Y. *Carbon* **62,** 455–464 (2013).
11. Parmentier, J., Gaslain, F. O. M., Ersen, O., Centeno, T. A. & Solovyov, L. A. Structure and sorption properties of a zeolite-templated carbon with the EMT structure type. *Langmuir* **30,** 297–307 (2014).
12. Kyotani, T., Ma, Z. & Tomita, A. Template synthesis of novel porous carbons using various types of zeolites. *Carbon* **41,** 1451–1459 (2003).
13. Yang, Z., Xia, Y. & Mokaya, R. Enhanced hydrogen storage capacity of high surface area zeolite-like carbon materials. *J. Am. Chem. Soc.* **129,** 1673–1679 (2007).
14. Yang, Z., Xia, Y., Sun, X. & Mokaya, R. Preparation and hydrogen storage properties of zeolite-templated carbon materials nanocast via chemical vapor deposition: effect of the zeolite template and nitrogen doping. *J. Phys. Chem. B* **110,** 18424–18431 (2006).
15. Zhou, J. *et al.* Effect of cation nature of zeolite on carbon replicas and their electrochemical capacitance. *Electrochim. Acta* **89,** 763–770 (2013).
16. Corma, A. From microporous to mesoporous molecular sieve materials and their use in catalysis. *Chem. Rev.* **97,** 2373–2420 (1997).
17. Davis, M. E. & Lobo, R. F. Zeolite and molecular sieve synthesis. *Chem. Mater.* **4,** 756–768 (1992).
18. Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417,** 813–821 (2002).
19. Corma, A. State of the art and future challenges of zeolites as catalysts. *J. Catal.* **216,** 298–312 (2003).
20. Kyotani, T., Nagai, T., Inoue, S. & Tomita, A. Formation of new type of porous carbon by carbonization in zeolite nanochannels. *Chem. Mater.* **9,** 609–615 (1997).
21. Johnson, S. A., Brigham, E. S., Ollivier, P. J. & Mallouk, T. E. Effect of micropore topology on the structure and properties of zeolite polymer replicas. *Chem. Mater.* **9,** 2448–2458 (1997).
22. Guisnet, M. & Magnoux, P. Organic chemistry of coke formation. *Appl. Catal. A Gen.* **212,** 83–96 (2001).
23. Deschamps, M. *et al.* Exploring electrolyte organization in supercapacitor electrodes with solid-state NMR. *Nature Mater.* **12,** 351–358 (2013).
24. Jun, S. *et al.* Synthesis of new, nanoporous carbon with hexagonally ordered mesostructured. *J. Am. Chem. Soc.* **122,** 10712–10713 (2000).
25. Liu, X., Giordano, C. & Antonietti, M. A facile molten-salt route to graphene synthesis. *Small* **10,** 193–200 (2014).
26. Ning, G. *et al.* Gram-scale synthesis of nanomesh graphene with high surface area and its application in supercapacitor electrodes. *Chem. Commun.* **47,** 5976–5978 (2011).

© 2016 Macmillan Publishers Limited. All rights reserved

27. Chakraborty, B., Modak, P. & Banerjee, S. Hydrogen storage in yttrium-decorated single walled carbon nanotube. *J. Phys. Chem.* **116,** 22502–22508 (2012).
28. Yoon, M. *et al.* Calcium as the superior coating metal in functionalization of carbon fullerenes for high-capacity hydrogen storage. *Phys. Rev. Lett.* **100,** 206806 (2008).
29. Odkhuu, D. *et al.* Negatively curved carbon as the anode for lithium ion batteries. *Carbon* **66,** 39–47 (2014).
30. Zhai, Y., Zhu, Z. & Dong, S. Carbon-based nanostructures for advanced catalysis. *ChemCatChem* **7,** 2806–2815 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by IBS-R004-D1. The authors thank D. Ahn at the Pohang Accelerator Laboratory (PLS) for discussions on powder XRD measurements. X-ray crystallography was carried out with help from D. Moon at PLS and H. J. Lee at Korea Basic Science Institute.

Author Contributions R.R. selected metal-ion catalysts intuitively, initiated single-crystal investigation, and led the project. K.K. led the synthesis and characterization work, with T.L. and Y.K. Y.S. carried out NMR measurements. J.S. and J.K.P. carried out electrochemical analysis. H.L. and J.Y.P. analysed the electrical conductivity of the carbon product. S.J.C. and T.L. carried out the X-ray crystallography. H.I. investigated the mechanism of carbon formation. R.R. and K.K. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.R. (rryoo@kaist.ac.kr).

Reviewer Information *Nature* thanks P. de Jongh, L. Solovyov and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Preparation of zeolite templates. Zeolite beta (Si/Al ratio = 12.5) was purchased from Zeochem, and MFI (Si/Al = 11.5) from Zeolyst. Other zeolites were synthesized according to literature procedures[31,32]. For X-ray crystallography, we synthesized zeolite FAU with large single-crystal morphology[33]. Ion exchange was performed with aqueous solution of metal salts.

Synthesis of carbon materials. Zeolite was heated to 600 °C under a dry N_2 flow, using a vertically placed, fused quartz reactor equipped with a fritted disk. We passed a mixture of ethylene gas, N_2 and steam through the zeolite bed at 600 °C. The gas flow was switched to dry N_2 when carbon deposition was completed. Then, we increased the temperature to 850 °C, and maintained it there for 2 hours. The resulting product was slurried in a 0.3 M HF/0.15 M HCl solution, or alternatively in concentrated HCl followed by hot 2 M NaOH solution, to release carbon from template. The HF-etched carbon products exhibited an oxygen/carbon ratio of 0.009 in molar ratio, while the carbon samples obtained with NaOH washing had an oxygen/carbon ratio of 0.10.

Collection of crystallographic data. A single crystal (about 35 μm in diameter; see Supplementary Fig. 8) of La FAU zeolite containing carbons, treated at 600 °C, was dehydrated by flowing high-purity N_2 gas at 600 °C for 2 hours, and then put under vacuum at 350 °C for 2 hours, to fully exclude moisture. A crystal was coated with a layer of Paratone oil (pre-dehydrated for 48 hours) in a glove box, in order to prevent moisture absorption during sample mounting and XRD measurements (see Supplementary Fig. 9 for experimental verification through a gravimetric measurement). The coated crystal was measured at 123 K over the range of $2\theta = 5°–149°$, using a Bruker D8-Venture diffractometer with graphite-monochromated CuKα radiation ($\lambda = 0.15418$ nm) and a Photon 100 CMOS detector, at the Korea Basic Science Institute (exposure time = 120 seconds per frame). The Bruker APEX2 program was used for data collection, and SAINT was used for cell refinement and reduction[34]. Absorption correction was applied using the SADABS program[35].

Derivation of electron-density maps for carbon. XRD data, collected from the zeolite–carbon composite crystal, were analysed by means of full-matrix least-squares calculations based on F^2 values with JANA2006 (ref. 36). We found the data to have an R_{int} (observed/all) value of 4.73/5.61 for 547/819 reflections, averaged from 3,996/7,829 reflections, with a redundancy and completeness of 9.559 and 99.9%, respectively. R_{int}, the merging error, is given by $R_{int} = \frac{\sum|F_0^2 - F_0^2(\text{mean})|}{\sum F_0^2}$, where F_0 is the experimental structure factor. We obtained an electron-density map by using a dual space method with a charge flipping algorithm[37]. We obtained the space group of $Fd\bar{3}m$, with an overall agreement factor of 2.85, from the electron-density map[38]. We discovered a total of eight atoms in an asymmetric unit, all in the zeolite framework. Refinement of the zeolite framework structure was started after assigning correct atom species using the full-matrix least-squares procedure. The value of maximum (change/s.u.), which is used as a convergence criterion, decreased from 0.05 to 0.01 during this initial refinement process (s.u. is the standard uncertainty). The values of R_1/wR_2 (which indicate the agreement between the crystallographic model and the experimental X-ray diffraction data) were 18.73/43.00 when Si and O atoms were taken into account for the framework.

For further refinement, including La and Na atoms, first, the scale factor of the zeolite crystal containing carbons was determined using the high-angle portion above $\sin\theta/\lambda = 0.25$ corresponding to $d \leq 0.2$ nm[39] (where d indicates the resolution, defined by the Breck equation, $\lambda = 2d\sin\theta$ This process yielded R_1/wR_2 values of 9.74/21.04 for 747 unique reflections. The R_1/wR_2 values decreased to 7.49/15.81 when occupancy factors for non-framework species (that is, Na, La, and O atoms bound to La) were refined. Second, anisotropic atomic displacement factors were used for all zeolite framework atoms and La atoms. Isotropic atomic displacement factors were used for non-framework atoms except La. This process further decreased R_1/wR_2 to 6.30/12.78. The resulting composition of the zeolite was $|La_{23.32}Na_{12.16}O_{12.12}|[T(Si,Al)O_2]_{192}$, which was consistent with that from the chemical analysis.

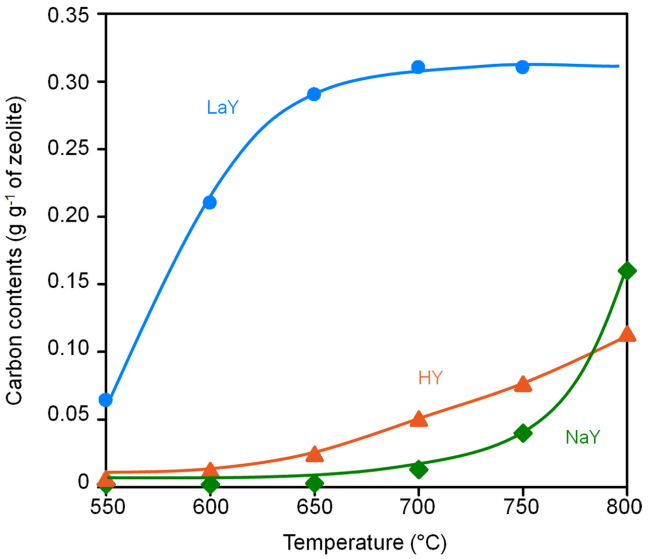
In this stage, we tried to refine intrapore species using all reflections including the low-angle portion (that is, 819 unique reflections). The scale factor, determined above, was fixed until most of the missing intrapore atoms were assigned from a difference Fourier method. In a first trial of the difference Fourier method, we found eight peaks. Two peaks were assigned as oxygen that was coordinated to La in the zeolite framework, and six peaks were assigned as carbon in the zeolite pore space. The refinement of occupancy factors and positions of the obtained eight missing peaks resulted in R_1/wR_2 decreasing to 7.07/17.91. The thermal displacement factor for all carbon atoms was set to a reasonable value, $U_{iso} = 0.08$ Å^2, according to ref. 40. By using the difference Fourier method again, we also found three carbon atoms. Refinement of the additional carbon and oxygen yielded R_1/wR_2 values of 6.66/15.72. Further inclusion of carbon atoms using the difference Fourier method did not improve the R factors. Thus, by this stage, it could be considered that most of the missing atoms were found. The obtained composition was $|C_{207.58}La_{23.52}Na_{14.27}O_{26.69}|\ [T(Si,Al)O_2]_{192}$. The carbon content is comparable to the empirical carbon content (258 atoms per unit cell) obtained from elemental analysis of the sample. The obtained composition was changed slightly by the subsequent additional refinement of the structural parameters with the scale factor.

In the last refinement, the occupancy factor of carbon atoms was fixed, and all carbon atoms were assumed to have the same atomic displacement parameter. The constraints listed in a CIF file (see Supplementary information) were automatically generated from the symmetry operation, based on this assumption. Refinement using these constraints was continued until R_1/wR_2 values reached 5.40/13.65 with 547 reflections for $I > 3\sigma(I)$ (where I is the reflection intensity). The largest difference peak was 0.50, and the deepest hole was −0.43 e Å^{-3}. The goodness-of-fitness index was 1.9. The final maximum (change/s.u.) was 0.0051. However, when all 819 reflections were taken into account, the R_1/wR_2 values were 7.89/14.36. Moreover, attempts to refine each carbon atomic displacement parameter independently, without using constraints, failed to give stable convergence. In this regard, the above refinement result using constraints may not yet provide an accurate structural solution. After removing the carbon atoms obtained from structure refinement with constraints, we visualized an electron-density map corresponding to the carbon structure using the difference Fourier method; this map was equivalent to the difference between the total electron-density map and an electron-density map that corresponds to the zeolite framework.

Characterization. We determined the carbon content in zeolites by thermogravimetry using a TGA Q50 (TA Instruments). We collected powder XRD data using a monochromated synchrotron X-ray at Beamline 9B of the Pohang Accelerator Laboratory. TEM images and SAED patterns were collected with a Titan E-TEM G2 (FEI) at 300 kV acceleration voltages, on a holey carbon grid (300 mesh) after supporting with ethanol dispersion. SEM images were taken with a Verios 460 (FEI) at a landing voltage of 1 kV in deceleration mode (stage bias voltage: 5 kV). ^{13}C NMR spectra were acquired with magic-angle spinning using Brucker Avance III HD 400WB. Raman spectra were recorded on a Horiba Jobin Yvon ARAMIS spectrometer with a laser excitation wavelength of 514 nm. Electrical conductance was measured using an Agilent atomic force microscope 5500 in air, with a Pt/Ir-coated tip (PPP-EFM-50, Nanosensors).

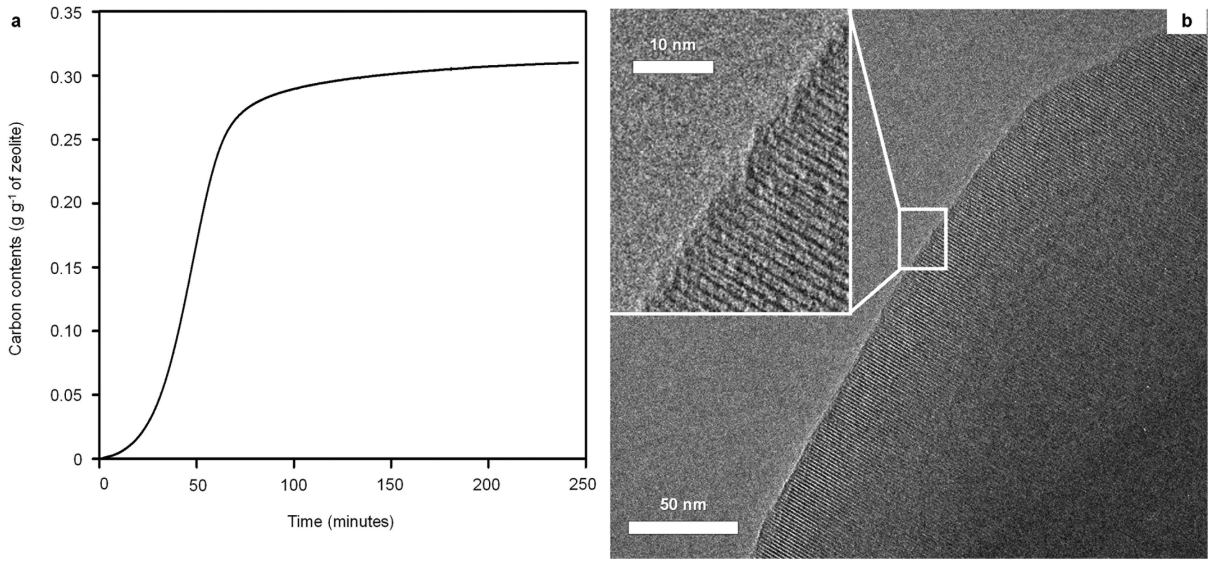
31. Breck, D. W. *Zeolite Molecular Sieves* (Wiley, 1974).
32. Delprato, F., Delmotte, L., Guth, J. L. & Huve, L. Synthesis of new silica-rich cubic and hexagonal faujasites using crown-ether-based supramolecules as templates. *Zeolites* **10,** 546–552 (1990).
33. Ferchiche, S., Warzywoda, J. & Sacco, A., Jr Direct synthesis of zeolite Y with large particle size. *Int. J. Inorg. Mater.* **3,** 773–780 (2001).
34. APEX2 and SAINT (Bruker AXS, 2014).
35. Sheldrick, G. M. *SADABS* (Univ. Göttingen, 2008).
36. Petříček, V., Dušek, M. & Palatinus, L. Crystallographic computing system JANA2006: general features. *Z. Kristallogr.* **229,** 345–352 (2014).
37. Palatinus, L. & Chapuis, G. SUPERFLIP—a computer program for the solution of crystal structures by charge flipping in arbitraray dimensions. *J. Appl. Cryst.* **40,** 786–790 (2007).
38. Palatinus, L. & van der Lee, A. Symmetry determination following structure solution in *P*1. *J. Appl. Cryst.* **41,** 975–984 (2008).
39. McCusker, L. B., Von Dreele, R. B. & Cox, D. E., Louër, D. & Scardi, P. Rietveld refinement guidelines. *J. Appl. Cryst.* **32,** 36–50 (1999).
40. Xie, D., McCusker, L. B. & Baerlocher, C. Structure of the borosilicate zeolite catalyst SSZ-82 solved using 2D-XPD charge flipping. *J. Am. Chem. Soc.* **133,** 20604–20610 (2011).
41. Biener, J. *et al.* Macroscopic 3D nanographene with dynamically tunable bulk properties. *Adv. Mater.* **24,** 5083–5087 (2012).

© 2016 Macmillan Publishers Limited. All rights reserved



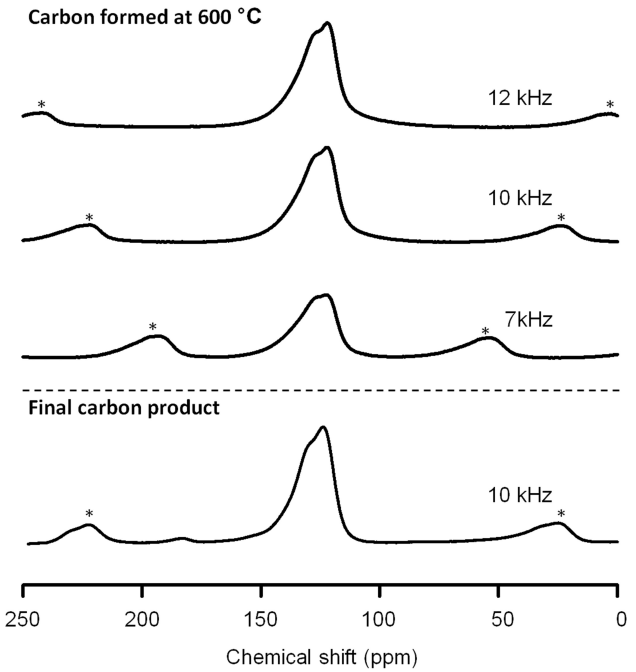
Extended Data Figure 1 | Carbon deposition in zeolite FAU plotted as a function of temperature, with various ion exchanges. The zeolites LaY, NaY and HY were heated to the temperatures indicated under a dry N_2 flow, using a vertically placed, fused quartz reactor equipped with a fritted disk. Subsequently, a mixture of ethylene gas, N_2 and steam was passed through the zeolite bed for 1 hour. The amount of carbon deposited at each temperature was measured by thermogravimetry. At 600 °C, the La^{3+}-ion-exchanged zeolite had been deposited with 20 times more carbon than had HY or NaY.

© 2016 Macmillan Publishers Limited. All rights reserved



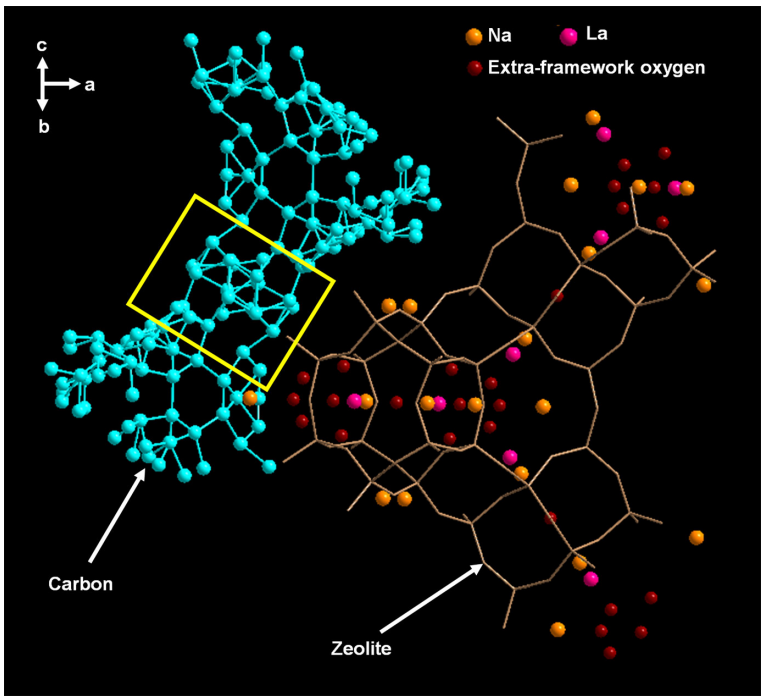
Extended Data Figure 2 | Carbon deposition in La^{3+}-ion-exchanged FAU zeolite at 600 °C. a, The amount of carbon was measured as a function of time, using thermogravimetric analysis equipment built in the carbon deposition rig. The plotted result indicates that the carbon content in LaY becomes saturated at $\sim$0.3 g g^{-1} of zeolite. **b**, A TEM image of LaY zeolite after 250 min of carbon deposition, showing apparently no carbon deposition on external surfaces.

© 2016 Macmillan Publishers Limited. All rights reserved



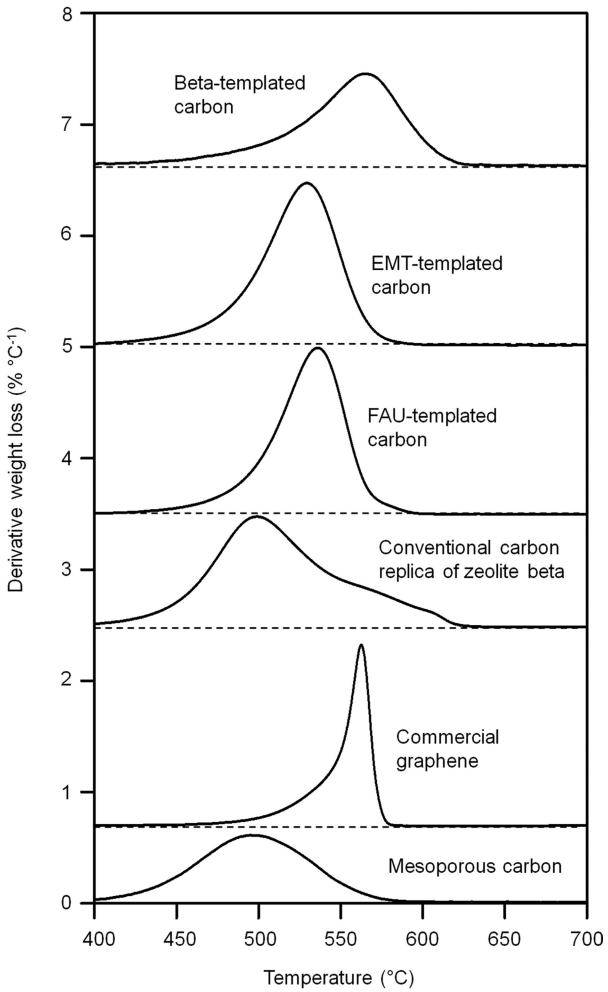
Extended Data Figure 3 | Magic-angle spinning solid-state ^{13}C NMR spectra of the carbon framework formed within zeolite LaY. The NMR spectra were recorded with various spinning rates on a Brucker Avance III HD 400WB NMR spectrometer operated at 100.61 MHz for ^{13}C. All spectra were obtained with a 4-μs pulse, a 10-s relaxation delay, and 1,000 acquisitions. Asterisks indicate spinning sidebands for a given spinning rate. The spectra for carbon obtained at 600 °C exhibit two peaks with chemical shift at 123 p.p.m. and 129 p.p.m. The peak at 123 p.p.m. can be assigned to six-membered-ring sp^2 carbon; the peak at 129 p.p.m. can be attributed to five- or seven-membered rings that have smaller C–C–C angles in the conjugated sp^2 carbon system[24]. No other peaks (assignable to sp^3 or sp carbons) were detected in the NMR spectra of sample prepared with 99% ^{13}C-isotope-enriched ethylene. The final carbon product, liberated from zeolite after heat treatment at 850 °C, has an additional weak peak at around 180 p.p.m., corresponding to oxygen functional groups.

© 2016 Macmillan Publishers Limited. All rights reserved



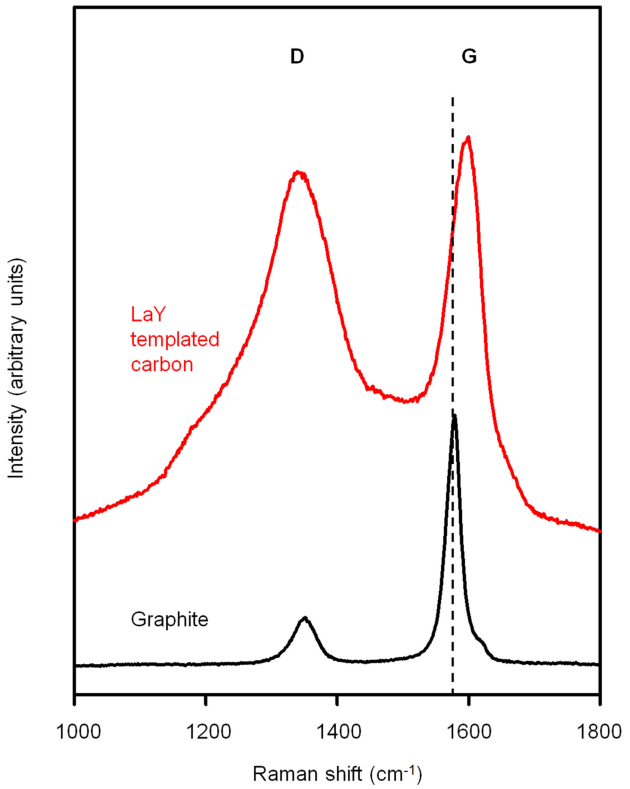
Extended Data Figure 4 | X-ray crystallographic analysis of the carbon structure formed in a single crystal of La^{3+}-ion-exchanged zeolite FAU. Carbon atomic positions were determined through least-square refinement of the distances, using a difference Fourier method (see Methods for details). To cope with a complex system having high static disorder of atomic positions, we assumed that all carbon atoms had the same thermal parameter in the refinement procedure. The refinement result indicates that atomic positions in pore necks (yellow rectangle) have high static disorders over a zeolite crystal. That is, the determined positions can be regarded as overlapped carbon positions over many identical pore necks. This result, using constraints, may not yet provide an accurate structural solution.

© 2016 Macmillan Publishers Limited. All rights reserved



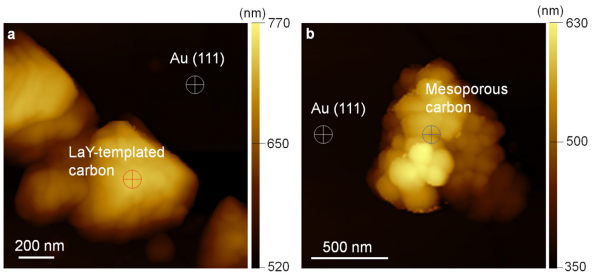
Extended Data Figure 5 | Thermal stability of carbon samples. The top three curves are derivative thermogravimetric curves for the carbons synthesized using different lanthanum-ion-exchanged zeolites. Thermogravimetry was carried out by increasing the temperature to 700 °C, with a ramping rate of 3 °C min^{-1}, under air flow (60 ml min^{-1}). We compared these thermogravimetric data with the results obtained using the mesoporous carbon CMK-3 (which has an amorphous structure), a commercial graphene product (purchased from Graphene Laboratories Inc.), and a beta-zeolite-templated carbon sample that was prepared following a two-step carbonization method[13] (bottom three curves). These data indicate that carbon samples obtained from lanthanum-ion-exchanged zeolites can have distinctively high thermal stability, compared with amorphous carbons. Notably, the beta-templated carbon exhibited high thermal stability in air, like graphene.

© 2016 Macmillan Publishers Limited. All rights reserved



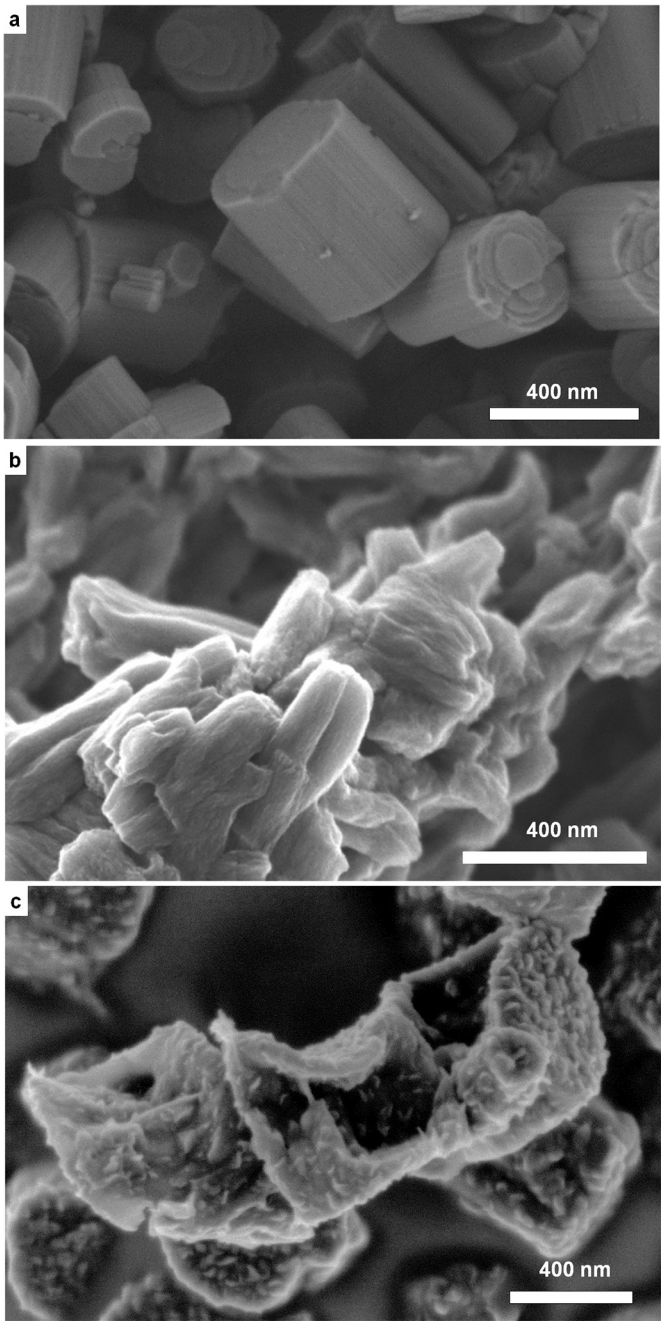
Extended Data Figure 6 | Raman spectra of LaY-templated carbon and graphite. The spectra were recorded on a Horiba Jobin Yvon ARAMIS spectrometer with a laser excitation wavelength of 514 nm. The G- and D-bands are located at 1,598 cm^{-1} and 1,341 cm^{-1}, respectively. The G-band of LaY-templated carbon appears at a higher wavenumber than that of graphite; such a strong upshift indicates nanosized single graphene layers[26]. The broad D-band is attributed to bond disorder, for instance because of the presence of five- or seven-membered carbon rings in the curved carbon structure[41].

© 2016 Macmillan Publishers Limited. All rights reserved



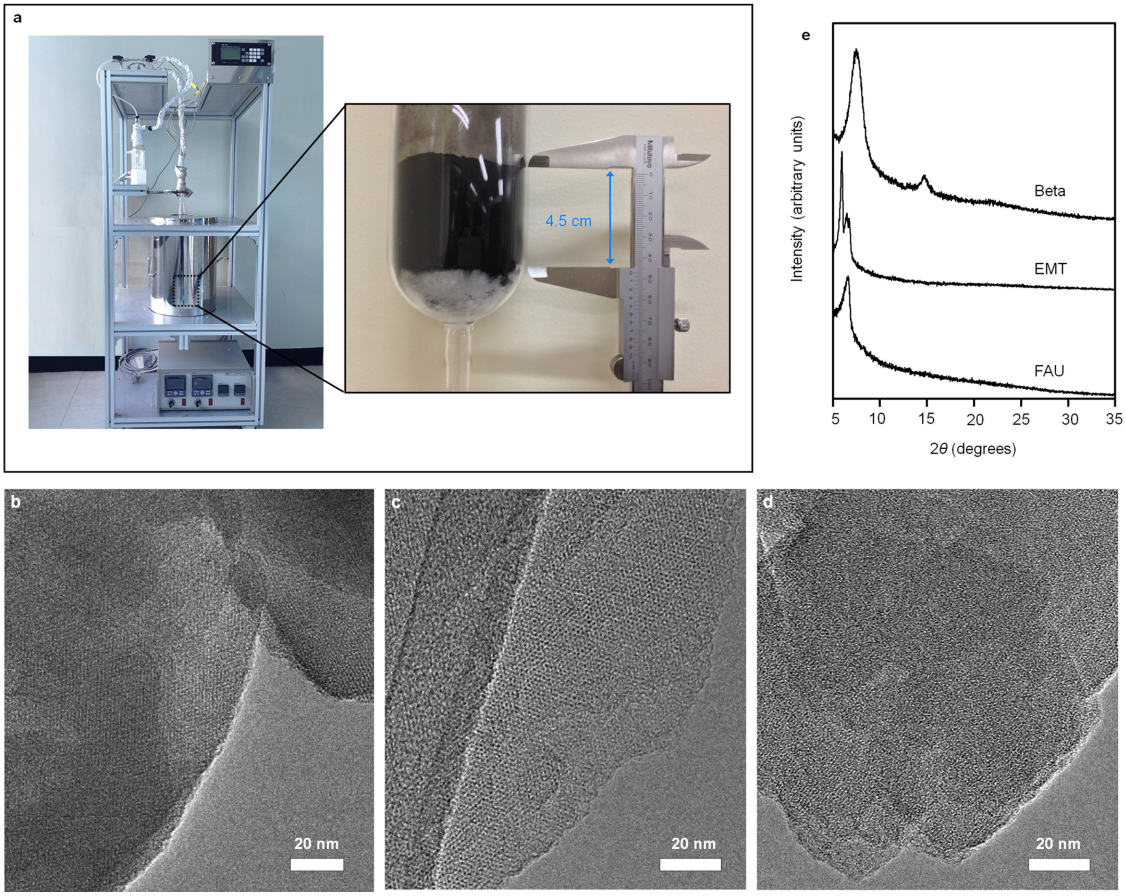
Extended Data Figure 7 | Topographical images of LaY-templated carbon and CMK-3 mesoporous carbon on an Au (111) substrate. **a**, LaY-templated carbon; **b**, CMK-3 mesoporous carbon. The current–voltage curves shown in Fig. 2f were measured on the cross-marked areas. The images were taken using an Agilent 5500 atomic force microscope in air, using a Pt/Ir-coated tip (see Methods).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | Effect of ion exchange on carbon synthesis using the 1D-channel LTL zeolite. a, SEM image of LTL zeolite. **b**, SEM image of carbon liberated from La^{3+}-ion-exchanged LTL zeolite. **c**, SEM image of carbon from H^{+}-ion-exchanged LTL zeolite. For carbon synthesis using La^{3+}-ion-exchanged zeolite, acetylene gas was used as the carbon source at 500 °C; the remainder of the protocol is as described in the Methods. For the H^{+}-ion-exchanged zeolite, carbon deposition was tested at various temperatures between 500 °C and 700 °C. However, synthesis using this H^{+} zeolite failed.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Scaling up the carbon-deposition process. **a**, Photograph of the carbonization rig for large-batch synthesis; inset, the plug-flow reactor filled with a thick bed of carbon–zeolite composite (about 40 g). From this apparatus, we could obtain about 10 g of batch carbon products in a single preparation. **b**–**d**, TEM images of the carbon products synthesized from zeolites FAU (**b**), EMT (**c**) and beta (**d**). **e**, XRD patterns of the carbons, confirming their highly ordered structure. These results indicate that the product quality from the 10-g batch synthesis is the same as that from the 0.15-g batch.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Data collection and refinement statistics for X-ray diffraction analysis

Formula	$\|C_{208.03}La_{23.52}Na_{15.50}O_{27.27}\|[T(Si,Al)O_2]_{192}$
Formula weight	18094.3
Temperature	123(2) K
Wavelength	Cu Kα (λ=0.15418 nm)
Crystal system	Cubic
Space group	$Fd\bar{3}m$
Unit cell dimensions	a = 25.0433 Å
Volume	15706.33 $Å^3$
Z	1
Density (calculated)	1.913 Mg/m^3
Absorption coefficient	16.682 mm^{-1}
F(000)	8737
Crystal size	0.035 x 0.035 x 0.035 mm^3
2θ range in data collection	149°
Index range	$-27 \le h \le 30, -15 \le k \le 31, -19 \le l \le 25$
Reflections collected	7829 [R_{int}(obs/all) = 4.73/5.61]
Completeness to theta = 74.5°	99.9%
Independent reflections (obs/all)	547/819
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	819 / 0 / 71
Goodness-of-fit on F^2	1.90
Final R indices [I>3sigma(I)]	R_1 = 0.0540, wR_2 =0.1365
R indices (all data)	R_1 = 0.0789, wR_2 =0.1436
Largest diff. peak and hole	0.50 and -0.43 e·$Å^{-3}$

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18010

Design of a hyperstable 60-subunit protein icosahedron

Yang Hsia[1,2,3]*, Jacob B. Bale[1,2,4]*, Shane Gonen[1,2,3,5], Dan Shi[5], William Sheffler[1,2], Kimberly K. Fong[1], Una Nattermann[1,2,3], Chunfu Xu[1,2], Po-Ssu Huang[1,2], Rashmi Ravichandran[1,2], Sue Yi[1,2], Trisha N. Davis[1], Tamir Gonen[5], Neil P. King[1,2] & David Baker[1,2,6]

The icosahedron is the largest of the Platonic solids, and icosahedral protein structures are widely used in biological systems for packaging and transport[1,2]. There has been considerable interest in repurposing such structures[3–5] for applications ranging from targeted delivery to multivalent immunogen presentation. The ability to design proteins that self-assemble into precisely specified, highly ordered icosahedral structures would open the door to a new generation of protein containers with properties custom-tailored to specific applications. Here we describe the computational design of a 25-nanometre icosahedral nanocage that self-assembles from trimeric protein building blocks. The designed protein was produced in *Escherichia coli*, and found by electron microscopy to assemble into a homogenous population of icosahedral particles nearly identical to the design model. The particles are stable in 6.7 molar guanidine hydrochloride at up to 80 degrees Celsius, and undergo extremely abrupt, but reversible, disassembly between 2 molar and 2.25 molar guanidinium thiocyanate. The icosahedron is robust to genetic fusions: one or two copies of green fluorescent protein (GFP) can be fused to each of the 60 subunits to create highly fluorescent 'standard candles' for use in light microscopy, and a designed protein pentamer can be placed in the centre of each of the 20 pentameric faces to modulate the size of the entrance/exit channels of the cage. Such robust and customizable nanocages should have considerable utility in targeted drug delivery[6], vaccine design[7] and synthetic biology[8].

Programming protein subunits to self-assemble into well defined complexes is a promising route to the custom design of macromolecular machines. Protein assemblies have been engineered using metals[9,10], disulfide bonds[11–14], genetic fusions[12,15–17], and ideal helix–helix interactions[11,13,16], but these approaches have generally yielded polydisperse or unanticipated products. Recently, symmetric modelling coupled with computational protein–protein interface design has accurately generated protein assemblies with tetrahedral and octahedral symmetry[18,19], but these relatively small (<16 nm diameter) nanocages have limited use for packaging or delivery applications because they have little internal volume.

Icosahedral point group symmetry contains two-, three-, and five-fold axes of rotation (Fig. 1a). To generate novel icosahedral protein assemblies, trimeric protein scaffolds of known structure were arranged with icosahedral symmetry (the three-fold axes of the trimers aligned with the three-fold axes of an icosahedron) and the two remaining degrees of freedom—the distance r from the icosahedron centre to the centre of mass of each trimer, and the angle ω of rotation of each trimer about its axis—were optimized for close packing without steric clashes (Fig. 1b, c). The amino acid sequences at the newly formed interfaces between the trimer building blocks were then optimized using RosettaDesign[20,21], and 17 designs were selected for experimental characterization on the basis of properties of the designed interface, including shape complementarity[22], predicted binding energy, and the number of buried unsatisfied hydrogen-bond donors and acceptors (see Methods).

Genes encoding the designs were assembled from oligonucleotides and cloned into the pET29b+ vector for expression in *E. coli*. Most of the designs were found in the insoluble fraction upon cell lysis; of the three soluble designs, two (both based on a KDPG aldolase[23,24]) showed substantial shifts in migration relative to the wild-type scaffold when analysed by native (non-denaturing) polyacrylamide gel electrophoresis (PAGE), suggesting higher-order assembly. We selected the one with fewer mutations, I3-01, for further analysis. Five substitutions (E26K, E33L, K61M, D187V and R190A) were made to generate the designed interface between trimers (Fig. 1d; the amino acid sequences are provided in the Supplementary Information).

I3-01 was purified using immobilized metal affinity and size exclusion chromatography (SEC), yielding a single peak with an apparent molecular weight much larger than that of the wild-type trimeric protein and consistent with the expected elution volume for the 60-subunit assembly (Fig. 1e). A mutant bearing a leucine-to-arginine substitution predicted to disrupt the designed interface eliminated the high-molecular-weight species and returned the elution volume to that of the wild-type scaffold (Fig. 1e). Dynamic light scattering (DLS) measurements of I3-01 showed a monodisperse population of particles with a hydrodynamic radius of 14 nm, consistent with the design model (Fig. 1f). No disassembly to the trimeric building block was observed at 80 °C or, remarkably, in 6.7 M guanidine hydrochloride (GuHCl) (Extended Data Fig. 1). This hyperstability is a property of both the trimeric scaffold from which I3-01 was derived and of the designed interface: both are completely resistant to GuHCl denaturation. An exceptionally sharp disassociation into the constituent trimers was observed between 2 M and 2.25 M guanidinium thiocyanate (GITC): at 2 M the dominant species is the icosahedron, while at 2.25 M only the trimeric building block is observed (Fig. 1g, Extended Data Fig. 2). Importantly for cargo packaging applications, the disassociation is fully reversible: the hydrodynamic radius of particles formed by diluting disassembled protein in 3 M GITC down to 1 M GITC is identical to those originally produced in *E. coli* (Fig. 1h).

We investigated the structure of I3-01 using cryo-electron microscopy (cryo-EM). The individual particles in large fields of view were homogenous in size and shape (Fig. 2a), and in class averages from 6,461 particles, the three projections along the symmetry axes and the overall icosahedral architecture are clearly discernible (Fig. 2b, c). A three-dimensional model calculated from the cryo-EM data matches

[1]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [2]Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. [3]Graduate Program in Biological Physics, Structure and Design, University of Washington, Seattle, Washington 98195, USA. [4]Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, Washington 98195, USA. [5]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA. [6]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved

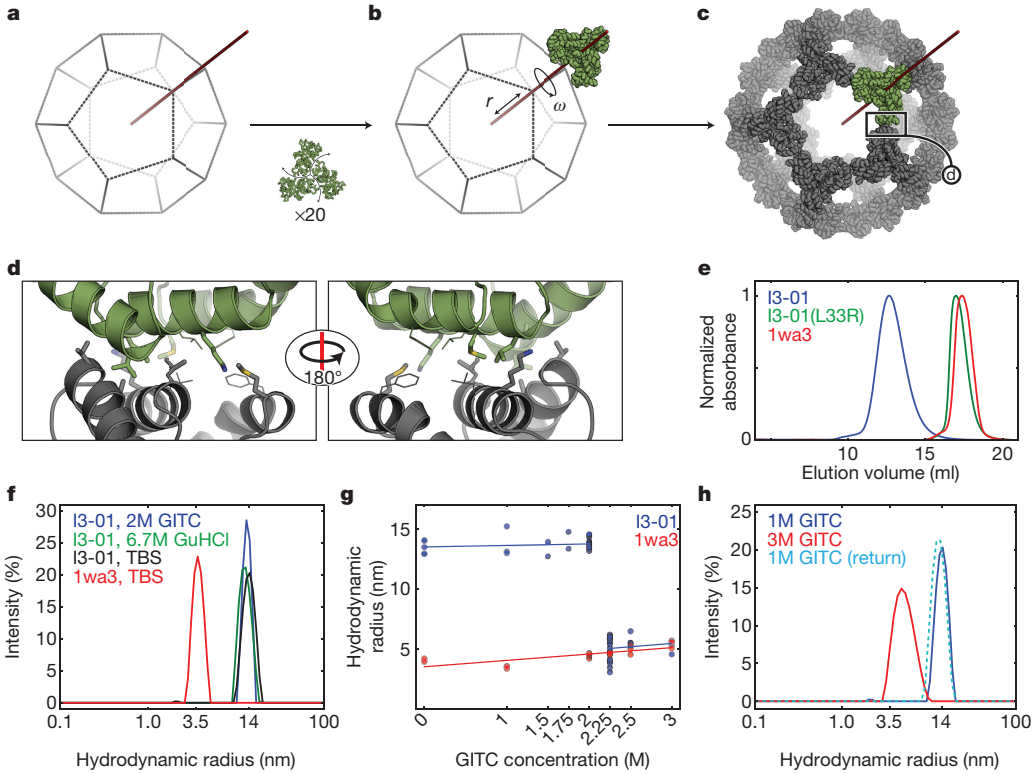


Figure 1 | Design methodology and biochemical characterization. **a**, **b**, Icosahedral three-fold axis in red and aligned trimeric building block in green. **c**, Optimization of r and ω yields closely opposed interfaces between subunits. **d**, Sequence design yields low-energy interfaces; in the I3-01 case, composed of five designed residues (thick representations) and two native residues (thin representations). **e**, I3-01 appears larger by SEC than the similarly sized I3-01(L33R) and wild-type trimer (1wa3). **f**, DLS measurement of hydrodynamic radius (note logarithmic scale in **f** and **h**) of 1wa3 (3.5 nm) and I3-01 (14 nm). I3-01 remains assembled in 6.7 M GuHCl and in 2 M GITC. **g**, Extremely sharp disassociation to trimeric building blocks at 2.25 M GITC. Data points represent independent measurements. **h**, I3-01 icosahedron disassembles into the trimeric building blocks at 3 M GITC, and reassembles following dilution to 1 M.

the I3-01 design model very well with a correlation coefficient of 0.92 at 20 Å and 1.5σ (Fig. 2d, e), clearly indicating that I3-01 forms the designed structure: an icosahedron with a diameter of 25 nm and an interior volume of approximately 3,000 nm^3, values that are within the range of those observed in small viral capsids[25].

To probe the robustness of I3-01 to genetic fusion, we fused superfolder GFP (sfGFP)[26] to one or both termini of the monomeric subunit and produced the resulting proteins in *E. coli*. SEC analysis showed that the fusion proteins had hydrodynamic radii consistent with cage formation (Extended Data Fig. 3). Analysis of I3-01 with a carboxy

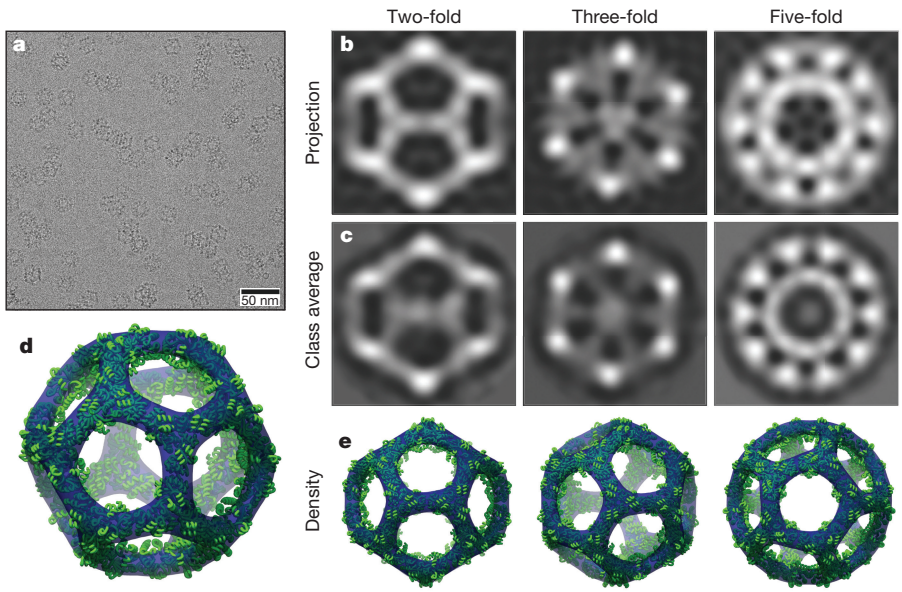


Figure 2 | Cryo-EM. a, Field-of-view cryo-EM micrograph showing homogeneous icosahedral particles in various orientations. **b**, Back-projections of I3-01 from the design model. **c**, Cryo-EM class averages closely match the design projections along all three symmetry axes. **d**, **e**, The calculated initial, unrefined density (blue, 3.22σ) closely matches the design model (green).

© 2016 Macmillan Publishers Limited. All rights reserved

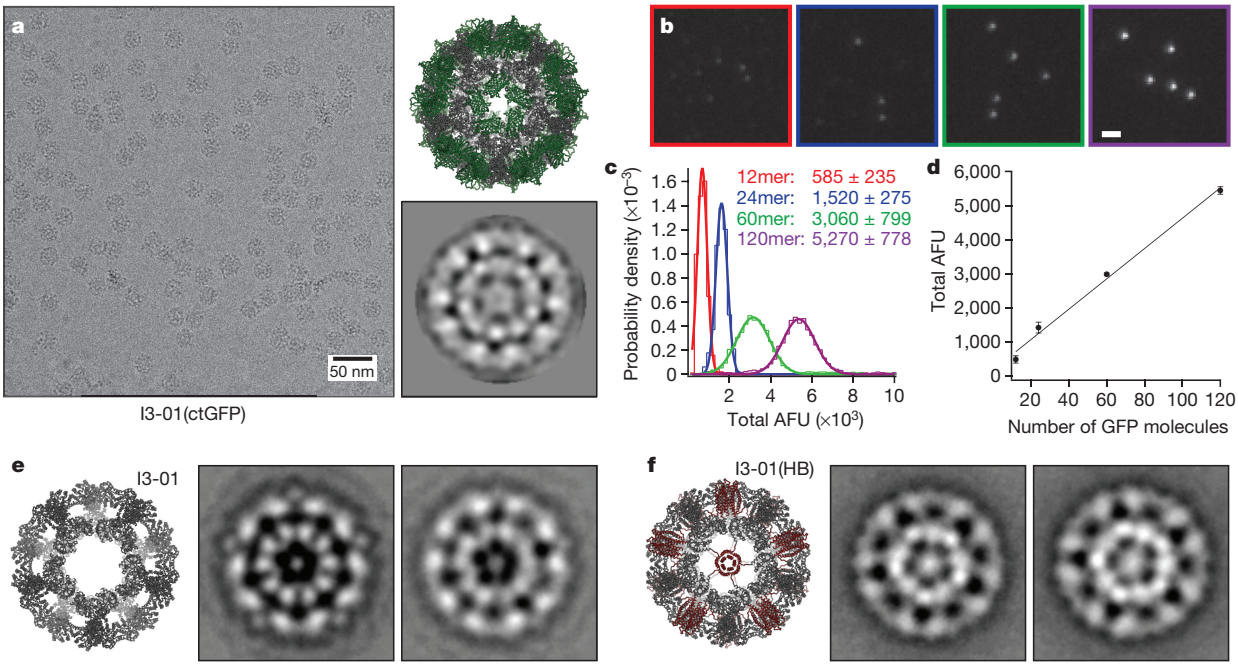


Figure 3 | Tuning nanocage structure and function with genetic fusions. **a**, The left panel shows a cryo-EM micrograph of I3-01(ctGFP); the top right panel shows a computational model with sfGFP in green; the bottom right panel shows the class average along the five-fold axis. **b**, Fluorescence microscopy fields of view. **c**, Fluorescence intensity histograms. AFU, arbitrary fluorescence units; ± standard deviation. **d**, Correlation between the mean fluorescence intensity and sfGFP copy number for nanoparticles with different numbers of fused sfGFP molecules. Error bars are s.e.m. ($n = 3$). **e**, **f**, Computational model and class averages along the five-fold axis of negatively stained I3-01 (**e**) and I3-01(HB) (**f**); the helical bundle is shown in red. Weak density in the centre of the pentameric faces in I3-01 may reflect randomly packaged material. There is clear density in the centre of the pentameric faces in the I3-01(HB) class averages consistent with the model.

(C)-terminal sfGFP fusion—called I3-01(ctGFP)—by cryo-EM revealed icosahedral particles with overall shapes very similar to those of the original design. Class averages of 13,297 particles revealed considerable internal density compared to the original I3-01 averages, consistent with computational models of the fusion complex (Fig. 3a). The I3-01 sfGFP fusions are robust to denaturation of the amino (N)- or C-terminal fused sfGFP in GuHCl; the particles remain assembled as GFP signal is lost[27] (Extended Data Fig. 4).

It is at present challenging to infer subunit copy number in GFP-tagged assemblies from their fluorescence intensity. What is needed are 'standard candles' with known fluorescent protein copy numbers that can be used to correlate fluorescence intensity to copy number. To complement the icosahedra with 60 and 120 copies of sfGFP described above, we fused sfGFP to one or both components of a previously described two-component tetrahedron (T33-21; ref. 19) to generate assemblies with 12 or 24 copies of sfGFP (Extended Data Fig. 3). Intensity histograms obtained for each of the sfGFP-nanocage constructs using widefield fluorescence microscopy were well fitted with Gaussians (Fig. 3b, c), and the mean fluorescence intensity for each cage was found to be linearly proportional ($r^2 = 0.9925$) to sfGFP copy number (Fig. 3d). The fluorescent properties of the particles were readily manipulated by substituting sfGFP with mTurquoise2 and sYFP2 (Extended Data Fig. 5). In addition to serving as genetically encoded, water-soluble fluorescent standard candles, the fluorescent protein cage fusions could be useful for correlative light and electron microscopy[28] since the icosahedral shape is quite distinctive.

In I3-01, the trimeric building blocks are aligned with the three-fold axes while the designed interface is along the icosahedral two-folds. To explore the possibility of symmetry-matched fusions to designed nanocages, we modelled a designed pentameric helical bundle[29] into the centre of the large 9-nm pore at the five-fold axis with a C-terminal linker; this fusion was named I3-01(HB). Negative-stain electron microscopy showed monodisperse particles of the expected size and symmetry; the incorporation of the pentamer does not interfere with icosahedron assembly. Particle averages showed a structure similar to that of the original icosahedron, with additional density at the centre of each five-fold axis, consistent with computational models of the fusion protein (Fig. 3e, f). The capability of incorporating symmetry-matched substructures into designed nanocages offers considerable flexibility and modularity; for example, pentamers filling otherwise open pentameric faces could control the release of cargo contained within the nanocage.

The designed I3-01 icosahedron is exceptionally stable, robust to genetic fusion, and has a considerably larger internal volume than previously designed nanocages with well defined and prespecified structures[14,17,19]. Enzymatic activity is retained in the assembled icosahedron (Extended Data Fig. 6), suggesting a route to custom nanoreactors. The ability to accurately design icosahedral protein structures opens the door to new approaches to vaccine generation and targeted drug delivery.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 January; accepted 13 April 2016.
Published online 15 June 2016.

1. Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. Origin of icosahedral symmetry in viruses. *Proc. Natl Acad. Sci. USA* **101,** 15556–15560 (2004).
2. Ritsert, K. *et al.* Studies on the lumazine synthase/riboflavin synthase complex of *Bacillus subtilis*: crystal structure analysis of reconstituted, icosahedral beta-subunit capsids with bound substrate analogue inhibitor at 2.4 Å resolution. *J. Mol. Biol.* **253,** 151–167 (1995).
3. Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* **22,** 485–491 (2011).
4. Roldão, A., Mellado, M. C. M., Castilho, L. R., Carrondo, M. J. T. & Alves, P. M. Virus-like particles in vaccine development. *Expert Rev. Vaccines* **9,** 1149–1176 (2010).
5. Effio, C. L. & Hubbuch, J. Next generation vaccines and vectors: designing downstream processes for recombinant protein-based virus-like particles. *Biotechnol. J.* **10,** 715–727 (2015).

© 2016 Macmillan Publishers Limited. All rights reserved

6. Ma, Y., Nolte, R. J. M. & Cornelissen, J. J. L. M. Virus-based nanocarriers for drug delivery. *Adv. Drug Deliv. Rev.* **64,** 811–825 (2012).
7. Smith, M. L. *et al.* Modified tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **348,** 475–488 (2006).
8. Bauler, P., Huber, G., Leyh, T. & McCammon, J. A. Channeling by proximity: the catalytic advantages of active site colocalization using Brownian dynamics. *J. Phys. Chem. Lett.* **1,** 1332–1335 (2010).
9. Brodin, J. D. *et al.* Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4,** 375–382 (2012).
10. Der, B. S. *et al.* Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **134,** 375–385 (2012).
11. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340,** 595–599 (2013).
12. Usui, K. *et al.* Nanoscale elongating control of the self-assembled protein filament with the cysteine-introduced building blocks. *Protein Sci.* **18,** 960–969 (2009).
13. Raman, S., Machaidze, G., Lustig, A., Aebi, U. & Burkhard, P. Structure-based design of peptides that self-assemble into regular polyhedral nanoparticles. *Nanomedicine* **2,** 95–102 (2006).
14. Raman, S. *et al.* Design of peptide nanoparticles using simple protein oligomerization domains. *Open Nanomed. J.* **2,** 15–26 (2009).
15. Sinclair, J. C., Davies, K. M., Vénien-Bryan, C. & Noble, M. E. M. Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6,** 558–562 (2011).
16. Boyle, A. L. *et al.* Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. *J. Am. Chem. Soc.* **134,** 15457–15467 (2012).
17. Lai, Y.-T. *et al.* Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6,** 1065–1071 (2014).
18. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336,** 1171–1174 (2012).
19. King, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510,** 103–108 (2014).
20. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487,** 545–574 (2011).
21. DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D. & André, I. Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* **6,** e20450 (2011).
22. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–950 (1993).
23. Griffiths, J. S. *et al.* Cloning, isolation and characterization of the *Thermotoga maritima* KDPG aldolase. *Bioorg. Med. Chem.* **10,** 545–550 (2002).
24. Fullerton, S. W. B. *et al.* Mechanism of the class I KDPG aldolase. *Bioorg. Med. Chem.* **14,** 3002–3010 (2006).
25. Perlmutter, J. D. & Hagan, M. F. Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66,** 217–239 (2015).
26. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).
27. Andrews, B. T., Schoenfish, A. R., Roy, M., Waldo, G. & Jennings, P. A. The rough energy landscape of superfolder GFP is linked to the chromophore. *J. Mol. Biol.* **373,** 476–490 (2007).
28. Cortese, K., Diaspro, A. & Tacchetti, C. Advanced correlative light/electron microscopy: current methods and new developments using Tokuyasu cryosections. *J. Histochem. Cytochem.* **57,** 1103–1112 (2009).
29. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346,** 481–485 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Howard Hughes Medical Institute (D.B. and T.G.), the JRC visitor programme (S.G.), the National Science Foundation CHE-1332907 (D.B.), a UW/Hutch CCSG Pilot Award NCI 5 P30 CA015704-41 (D.B. and N.P.K.), Takeda Pharmaceutical Company (N.P.K.), the Bill and Melinda Gates Foundation OPP1120319 (D.B. and N.P.K.), and the National Institute of Health P41 GM103533 (T.N.D.). Y.H. was supported in part by a NIH Molecular Biology Training Grant (T32GM008268). U.N. was supported in part by a PHS National Research Service Award (T32GM007270) from NIGMS. J.B.B. was supported in part by an NSF Graduate Research Fellowship (DGE-0718124). We thank the Janelia Research Campus Cryo-EM Facility and J. de la Cruz for their assistance with the Titan Krios.

Author Contributions J.B.B., N.P.K., and W.S. developed the computational design methodology. Y.H. and J.B.B. performed the design of the icosahedra. Y.H. performed all other unlisted experiments. S.G. and D.S. performed the cryo-EM experiments. K.K.F. performed the fluorescence microscopy experiments. U.N. performed the negative-stain electron microscopy experiments. C.X. provided the pentamer sequence for I3-01(HB). P.-S.H. created the computational methodology to model fusions to I3-01. R.R. produced I3-01(HB) proteins. S.Y. produced T33-21 sfGFP fusion proteins.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (dabaker@uw.edu).

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Computational design. Crystal structures of 300 trimers with resolution better than 2.5 Å and lacking long loops were selected from the Protein Data Bank (PDB) to use as building blocks. For each scaffold, 20 trimeric building blocks were arranged in icosahedral symmetry by aligning the three-fold rotational axis of each trimer with one of the three-fold icosahedral symmetry axes. While preserving symmetry, the building blocks were then docked together by enumeratively sampling their rotations (ω) about the three-fold symmetry axes and translating (r) them into contact along the aligned axes. Configurations in which backbone atoms from different building blocks were less than 3.5 Å apart were discarded. Non-clashing design models were ranked based on the number of pairs of β-carbons in adjacent subunits within 12 Å and further sampling was carried out around the top 208 docked configurations on a 0.5 Å by 0.2 Å grid. Symmetric RosettaDesign[20,21] calculations were then used to generate low-energy, symmetric hydrophobic interfaces, and the resulting designs were filtered based on shape complementarity[22] (sc), interface surface area (sasa), buried unsatisfied hydrogen bonds (uhb), and binding energy (ddg). Designed substitutions that did not substantially contribute to the interface were reverted to their original identities. All Rosetta scripts used are available upon request; the full 60-subunit design model of I3-01 is provided in the Supplementary Information.

Cloning, screening, and protein purification. Codon-optimized genes encoding the wild-type and the designed molecules were generated by recursive polymerase chain reaction (PCR) from sets of synthetic oligonucleotides (Integrated DNA Technologies). Five mutations were incorporated into I3-01: E26K, E33L, K61M, D187V and R190A. All genes were cloned into the pET29b+ plasmid with kanamycin resistance and expressed in BL21 Star (DE3) *E. coli* cells (Invitrogen) induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 37 °C. Cell lysis was accomplished in Tris-buffered saline (TBS; 50 mM Tris, 500 mM NaCl) with lysozyme (0.25 mg ml^{-1}) and sonication (Fisher Scientific) at 20 W for 5 min total 'on' time, using cycles of 10 s on, 10 s off.

For initial screening, all constructs were labelled with the CoA-488 fluorophore (NEB) by the addition of AcpS[30] (NEB) using an A1 peptide tag, allowing the solubility and assembly state of each design to be analysed using SDS–PAGE and native-PAGE (Bio-Rad), following procedures previously described[18]. All subsequent experiments were performed on either $(His)_6$-tagged protein or remained untagged.

After lysis and centrifugation at 20,000*g* for 30 min, the soluble fraction of $(His)_6$-tagged proteins were passed through 2 ml of nickel nitrilotriacetic acid agarose (Ni-NTA) (Qiagen), washed with 30 mM imidazole, and eluted with 500 mM imidazole. Pure proteins were collected after elution from a Superose 6 10/300 GL SEC column (GE Healthcare) at 9–11 ml, depending on the fusion variant.

For non-$(His)_6$-tagged proteins, cells were lysed as above, and the cleared lysates were treated with serial ammonium sulfate precipitation treatments (20%, 60% w/v). During each step, solid ammonium sulfate was added to the lysate to the desired percentage, and equilibrated at room temperature for 1 h. Ammonium sulfate precipitated protein was then collected by centrifugation at 20,000*g* for 30 min at 25 °C. After treatment at 60%, the pellet was then solubilized in TBS and heated at 80 °C for 10 min. The soluble fraction was then collected and further purified through SEC as described.

KDPG enzyme assay. The reaction was carried out in 25 mM HEPES, 20 mM NaCl buffer at pH 7.0 with the presence of NADH (0.1 mM), L-lactate dehydrogenase (LDH, 0.11 U μl^{-1}), and 2-keto-3-deoxy-6-phosphogluconate (KDPG, 1 mM) at 25 °C, based on previously described methods[23]. Native 1wa3, I3-01, or I3-01(K129A) was added at 0.02 μM final concentration to each well and immediately monitored for 339 nm ultraviolet absorbance over time.

Dynamic light scattering. Purified protein was measured using a DynaPro NanoStar (Wyatt) DLS setup. 0.5 mg ml^{-1} of I3-01 and 1wa3 were measured at 25 °C, then the temperature was ramped up to 90 °C, then ramped back down to 25 °C for temperature scans at 2 °C min^{-1}. Measurements were taken in the presence of TBS: 25 mM Tris, 500 mM NaCl; buffered GuHCl: 25 mM Tris, 500 mM NaCl, 1–6.7 M GuHCl; or buffered GITC: 25 mM Tris, 500 mM NaCl, 1–4 M GITC. Different concentrations of GITC equilibrated samples were achieved by combining stocks of 0 M and 4 M equilibrated solutions in different ratios while GuHCl equilibrated samples were equilibrated individually. Each sample was allowed to equilibrate in their respective buffer for at least 24 h before measurement. Re-annealing experiments were performed by diluting I3-01 equilibrated in 3 M GITC down to 1 M GITC final concentration (0.166 mg ml^{-1} protein). Data analysis was performed using DYNAMICS v7 (Wyatt), reporting regularization fits (with D10/D50/D90) except for temperature ramp experiments, where cumulant fits were used. The ~1 nm radius particle consistent with GITC buffer alone was disregarded for analysis, and monodispersity was assumed when peak polydispersity was below 15% (refs 31 and 32).

Negative-stain electron microscopy. 3 μl of purified I3-01 and I3-01(ctGFP) at 0.1 mg ml^{-1} were applied to glow discharged, carbon-coated 200-mesh copper grids (Ted Pella), washed with Milli-Q water and stained with 0.75% uranyl formate as described previously[33]. Grids were visualized for assembly validation and stability and subsequently optimized for cryo-EM data collection. Screening was performed on a 120 kV Tecnai Spirit T12 transmission electron microscope (FEI) with a bottom-mount TVIPS F416 CMOS 4k camera.

6 μl of purified I3-01 and I3-01(HB) at 0.05–0.1 mg ml^{-1} were applied to glow discharged, carbon-coated 400-mesh copper grids (Ted Pella), washed with Milli-Q water and stained with 0.75% uranyl formate. Grids were visualized for assembly validation and optimized for data collection. Screening and sample optimization was performed on a 100 kV Morgagni M268 transmission electron microscope (FEI) equipped with an Orius charge-coupled device (CCD) camera (Gatan). Data collection was performed on a 120 kV Tecnai G2 Spirit transmission electron microscope (FEI). All final images were recorded using an Ultrascan 4000 4 k × 4 k CCD camera (Gatan) at 52,000× magnification at the specimen level. Coordinates for 6,576 I3-01 and 4,131 I3-01(HB) unique particles were obtained for averaging using EMAN2[34]. Boxed particles were used to obtain two-dimensional class averages by refinement in EMAN2. Additional image analysis was performed using ImageJ[35].

Cryo-EM. 5 μl of purified untagged I3-01 and I3-01(ctGFP), diluted to ~0.1 mg ml^{-1} using TBS buffer (25 mM Tris pH 8.0, 150 mM NaCl) with an additional 2 mM dithiothreitol were applied to glow discharged 1.2/1.3 Quantifoil grids, blotted and plunged into liquid ethane using a Vitrobot (FEI). Screening and grid optimization was performed on a 200 kV TF20 transmission electron microscope (FEI) with a bottom-mount TVIPS F416 CMOS 4k camera. 4–6 s movies were recorded on a 300 kV Titan Krios (FEI) using a Gatan K2 direct detector at either 29,000× or 37,000× magnification at the specimen level at ~10 electrons per pixel per second.

Movies were motion-corrected using previously described methods[36]. Coordinates for 6,461 (I3-01) and 13,297 (I3-01(ctGFP)) unique particles were obtained for averaging using EMAN2[34]. Extracted frames of these particles were used to calculate class averages by refinement in IMAGIC[37] using multiple rounds of multivariate statistical analysis and multi-reference alignment. An initial density model was calculated based on the calculated averages using EMAN2[34] and the fitting of the model and correlation were calculated using UCSF Chimera[38]. Low-resolution (17–30 Å) volumes from the I3-01 design model were calculated using SPIDER[39] and inspected in UCSF Chimera[38]. Back-projection images were computed in SPIDER[39] on the low-resolution volumes and visualized using WEB[39]. The contrast of all micrographs was enhanced in Fiji[40].

Symmetrical linker modelling. RosettaRemodel[41] was used to model I3-01(ctGFP) and to generate linkers for I3-01(HB). For I3-01(ctGFP), I3-01 was held static while the linker was sampled via fragment insertion, placing the sfGFP molecules at the end of the linker. The overall model was sampled symmetrically with icosahedral symmetry.

For I3-01(HB), I3-01 was held static while linkers of different lengths (7–12 residues) were sampled via fragment insertion. The resulting placement of the helical bundle at the end of the linker was filtered with pentameric assembly constraints to determine linker lengths that could satisfy formation of the pentameric helical bundle. The shorter linkers that allowed unstrained helical assembly were selected for experimental testing. Example scripts are supplied in the Supplementary Information.

Fluorescence microscopy. Different constructs used for fluorescence microscopy were generated by genetically fusing sfGFP to the termini of nanocages. For T33-21, the sfGFP was fused to either the C terminus of the first component (12 sfGFP molecules), or the C terminus of both components (24 sfGFP molecules). For I3-01, the sfGFP was fused to either terminus of I3-01 (60 sfGFP molecules), or both termini of I3-01 (120 sfGFP molecules). For mTurquoise2 and SYFP2 versions, sfGFP was replaced with the sequence of the respective fluorescent protein bearing additional surface mutations identical to sfGFP[26].

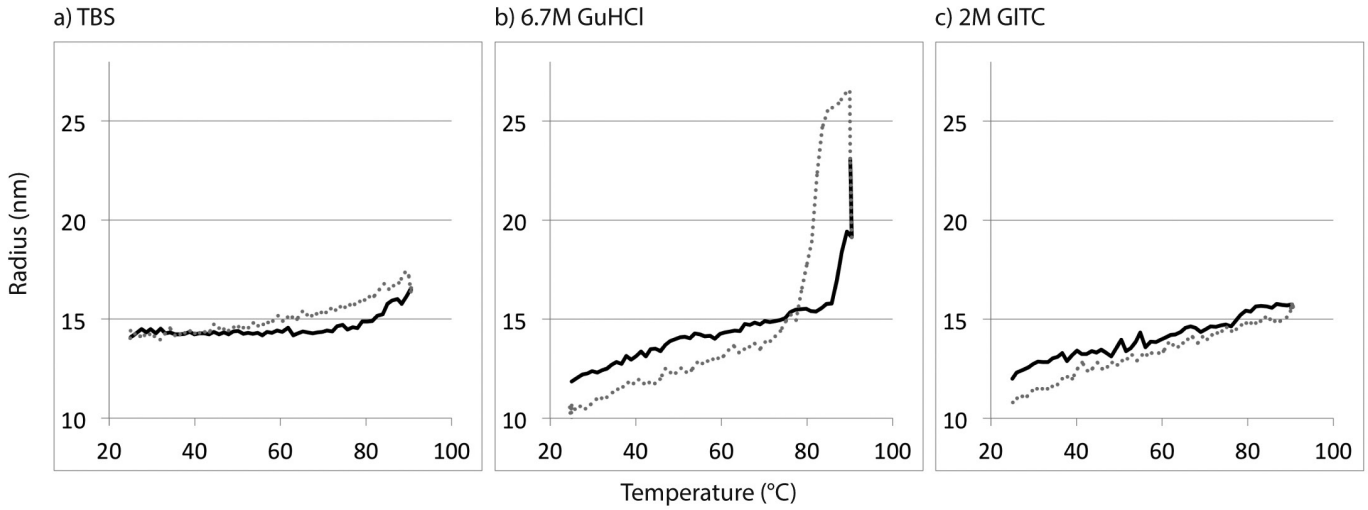
GFP nanocages were mounted on agarose pads for microscopy as previously described[42]. Images of the GFP nanocages were obtained using a DeltaVision system (Applied Precision) with an IX70 inverted microscope (Olympus), a U Plan Apo 100× objective (1.35 NA) and a CoolSnap HQ digital camera (Photometrics). GFP images were taken with a 0.4 s exposure, in a single focal plane, and binned 1 × 1.

The fluorescence intensities of GFP puncta were identified and quantified using custom Matlab programs as previously described[43]; programs are available upon request. Fluorescent intensity histograms of individual sfGFP-fused cages were fitted with Gaussian distributions, shown with mean total arbitrary fluorescence unit (AFU) intensity ± one standard deviation.

© 2016 Macmillan Publishers Limited. All rights reserved

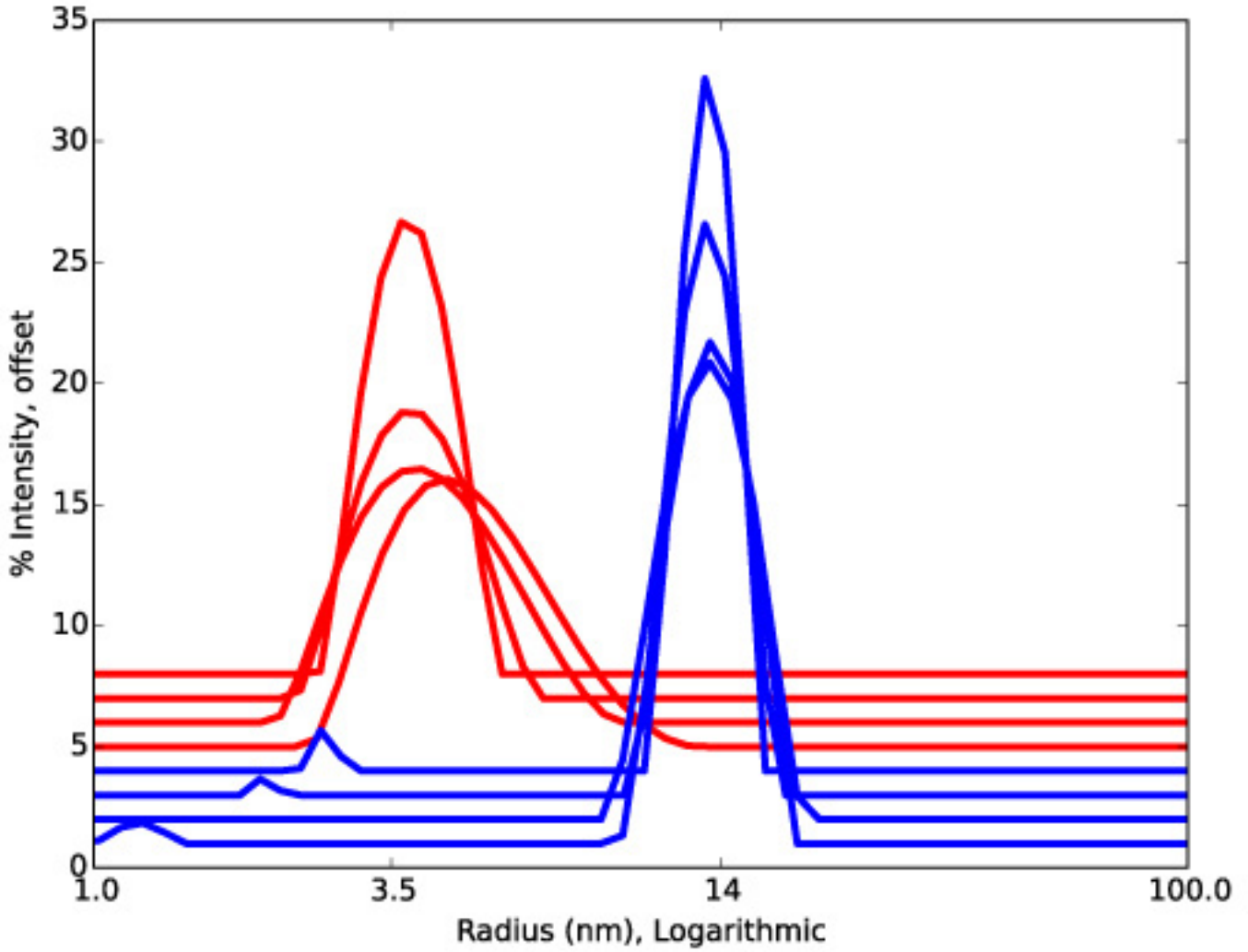
30. Zhou, Z. *et al.* Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. *ACS Chem. Biol.* **2,** 337–346 (2007).
31. Baalousha, M. & Lead, J. R. Nanoparticle dispersity in toxicology. *Nat. Nanotechnol.* **8,** 308–309 (2013).
32. Zulauf, M. & D'Arcy, A. Light scattering of proteins as a criterion for crystallization. *J. Cryst. Growth* **122,** 102–106 (1992).
33. Nannenga, B. L., Iadanza, M. G., Vollmar, B. S. & Gonen, T. in *Current Protocols in Protein Science* (eds Coligan, J. E., Dunn, B. M., Speicher, D. W. & Wingfield, P. T.) Ch. 17.15 (John Wiley & Sons, 2013).
34. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157,** 38–46 (2007).
35. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9,** 671–675 (2012).
36. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10,** 584–590 (2013).
37. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116,** 17–24 (1996).
38. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
39. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116,** 190–199 (1996).
40. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).
41. Huang, P.-S. *et al.* RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS One* **6,** e24109 (2011).
42. Muller, E. G. D. *et al.* The organization of the core proteins of the yeast spindle pole body. *Mol. Biol. Cell* **16,** 3341–3352 (2005).
43. Shimogawa, M. M., Wargacki, M. M., Muller, E. G. & Davis, T. N. Laterally attached kinetochores recruit the checkpoint protein Bub1, but satisfy the spindle checkpoint. *Cell Cycle* **9,** 3619–3628 (2010).

© 2016 Macmillan Publishers Limited. All rights reserved



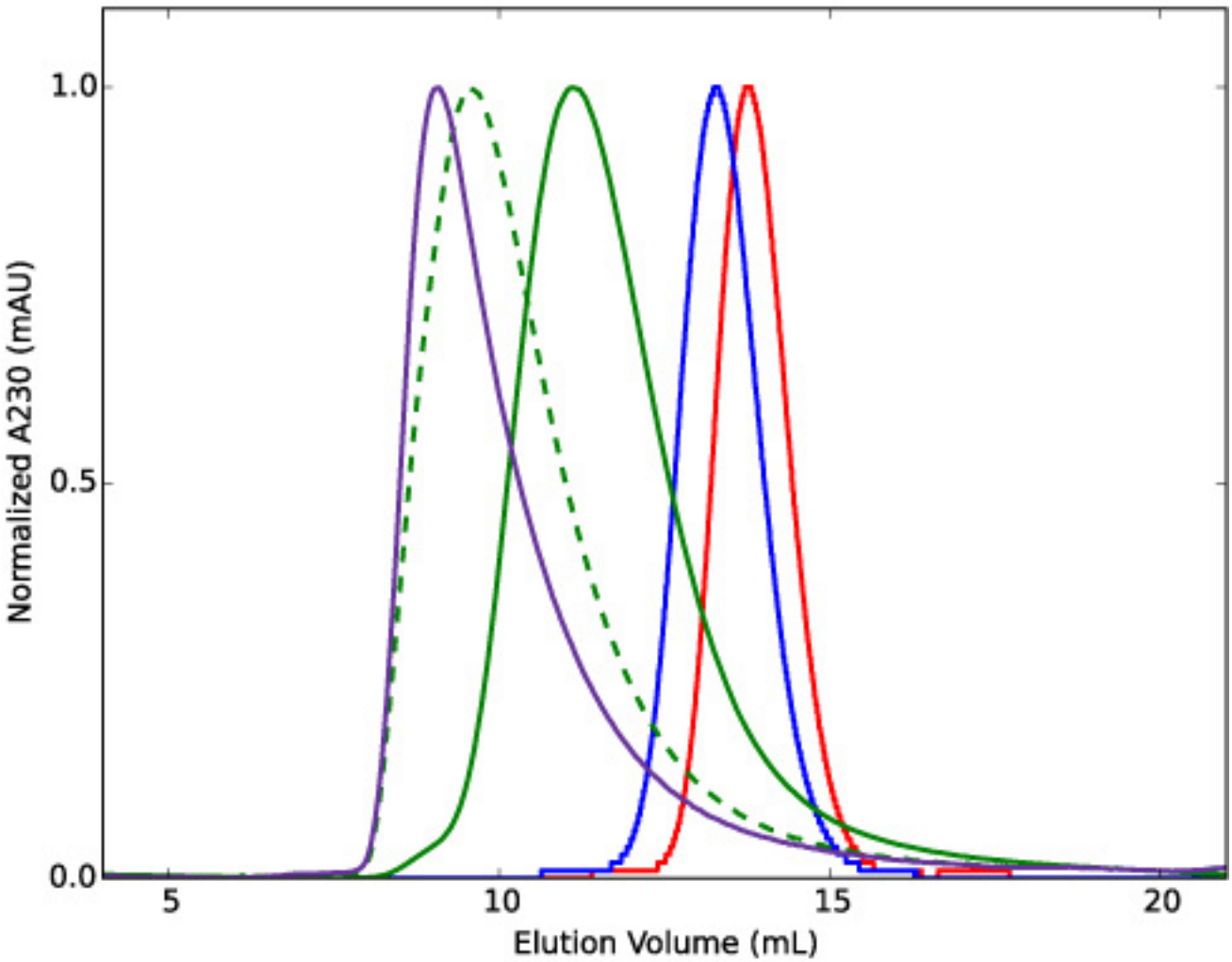
Extended Data Figure 1 | I3-01 tolerance to temperature. DLS measurements as I3-01 is subjected to heating to 90 °C (solid line), then cooling to 25 °C (dotted line) in TBS (**a**), 6.7 M GuHCl (**b**) and 2 M GITC (**c**). Under all three conditions, any indications of aggregation or increase in size due to temperature appear to be completely reversible.

© 2016 Macmillan Publishers Limited. All rights reserved



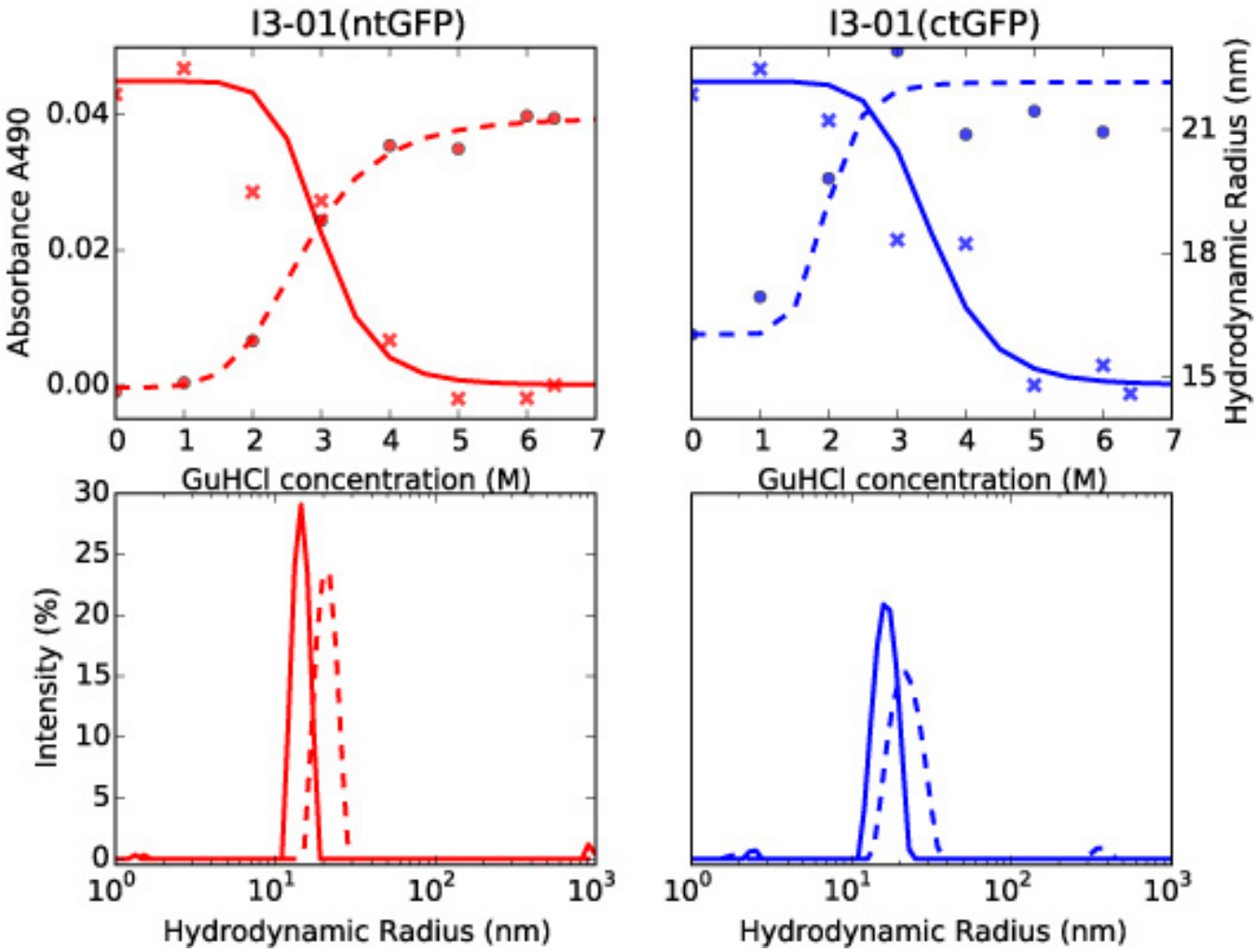
Extended Data Figure 2 | Reproducibility of I3-01 transition in 2 M to 2.25 M GITC. Four examples each of independent measurements at 2 M (blue) and 2.25 M (red) GITC using DLS show the reproducibility of the cage disassociation. Histograms are plotted offset by 1% intensity from each other for clarity.

© 2016 Macmillan Publishers Limited. All rights reserved



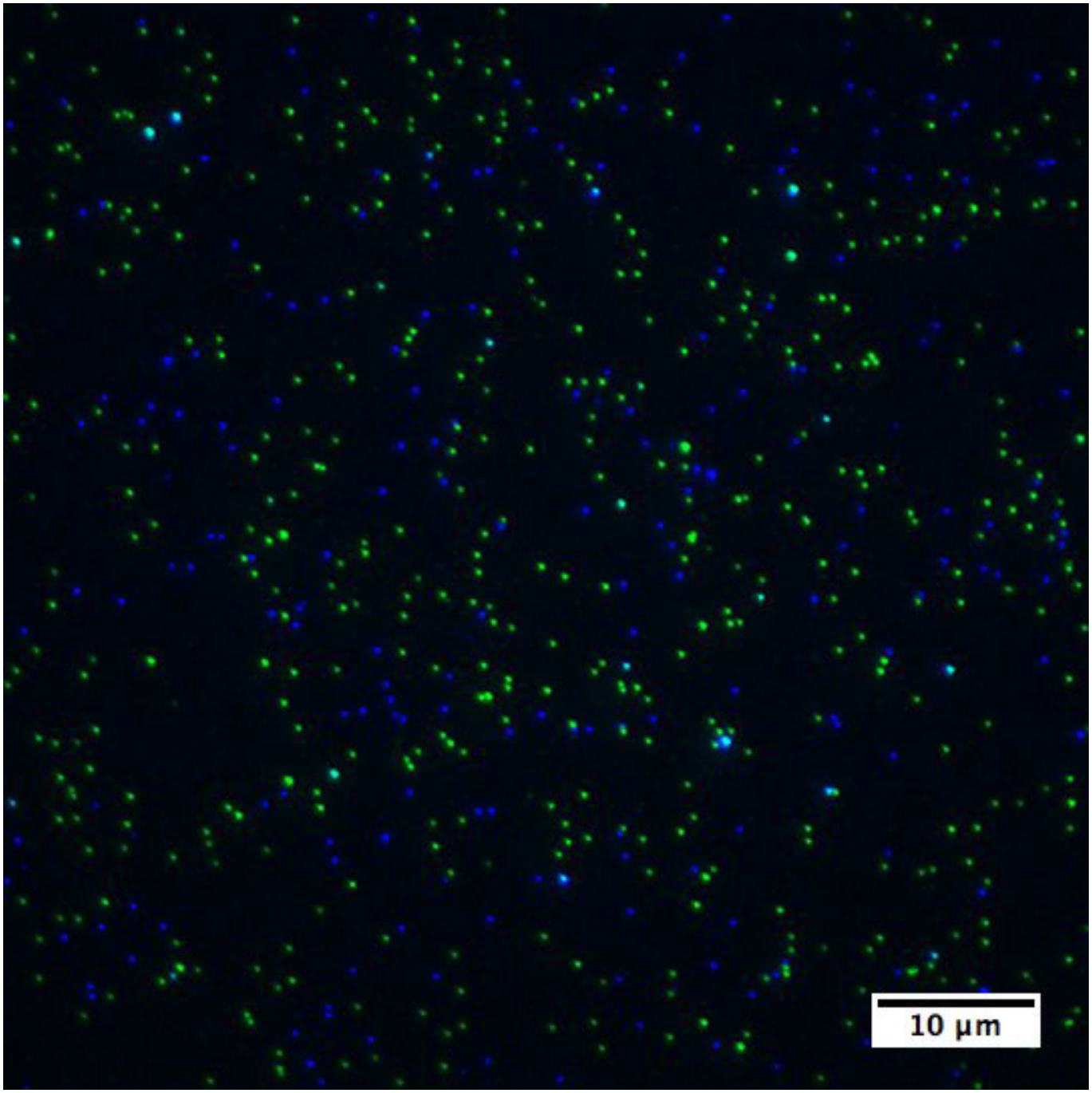
Extended Data Figure 3 | SEC of T33-21 and I3-01 fused with sfGFP. Size exclusion chromatography traces for T33-21 (12mer in red and 24mer in blue) and I3-01 (60mer in green and 120mer in purple) sfGFP fusions, display increased particle sizes with increasing copies of GFP, but retain monodispersed populations. The N-terminal fusion of sfGFP (dashed line) is expected to extend mostly outward from the icosahedron, thus greatly increasing the hydrodynamic radius while the C-terminal fusion is predicted to occupy the internal void space. A230, ultraviolet absorbance at 230 nm; mAU, milli-absorbance units.

© 2016 Macmillan Publishers Limited. All rights reserved



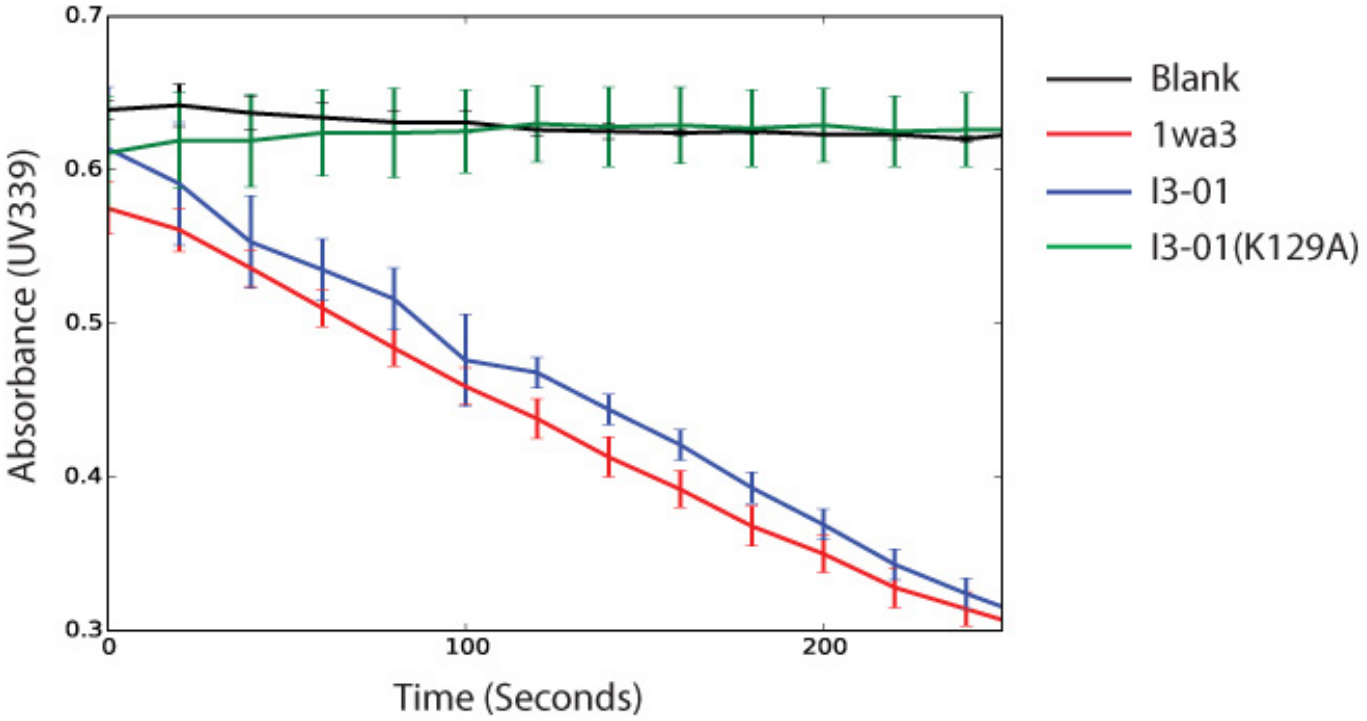
Extended Data Figure 4 | Tolerance of I3-01–sfGFP fusions to GuHCl. N-terminal (red) and C-terminal (blue) sfGFP fusions were equilibrated to 0–6.4 M GuHCl. Ultraviolet absorbance at 490 nm (A490) monitors the unfolding of sfGFP (top, solid line and crosses). DLS experiments (top, dotted line and dots) reveal as sfGFP unfolds, the hydrodynamic radius increases slightly, and then stabilizes. The bottom panels show that in 1 M GuHCl (solid line) and in 6 M GuHCl (dotted line), the icosahedral assemblies remain relatively monodisperse.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | I3-01 C-terminal fusions with other fluorescent proteins. Fluorescent proteins mTurquoise2 (in blue) or sYFP2 (in green) were fused to the C terminus of I3-01. The field of view using widefield fluorescence microscopy shows distinct signals of each type when the two types are mixed together.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 6 | I3-01 retains native enzyme activity. Coupled KDPG aldolase assay showing native-like enzymatic activity in I3-01. The K129A knockout shows no enzyme activity, similar to buffer alone. UV339, absorbance at 339 nm; error bars are standard deviation.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature17992

Subduction controls the distribution and fragmentation of Earth's tectonic plates

Claire Mallard[1], Nicolas Coltice[1,2], Maria Seton[3], R. Dietmar Müller[3] & Paul J. Tackley[4]

The theory of plate tectonics describes how the surface of Earth is split into an organized jigsaw of seven large plates[1] of similar sizes and a population of smaller plates whose areas follow a fractal distribution[2,3]. The reconstruction of global tectonics during the past 200 million years[4] suggests that this layout is probably a long-term feature of Earth, but the forces governing it are unknown. Previous studies[3,5,6], primarily based on the statistical properties of plate distributions, were unable to resolve how the size of the plates is determined by the properties of the lithosphere and the underlying mantle convection. Here we demonstrate that the plate layout of Earth is produced by a dynamic feedback between mantle convection and the strength of the lithosphere. Using three-dimensional spherical models of mantle convection that self-consistently produce the plate size–frequency distribution observed for Earth, we show that subduction geometry drives the tectonic fragmentation that generates plates. The spacing between the slabs controls the layout of large plates, and the stresses caused by the bending of trenches break plates into smaller fragments. Our results explain why the fast evolution in small back-arc plates[7,8] reflects the marked changes in plate motions during times of major reorganizations. Our study opens the way to using convection simulations with plate-like behaviour to unravel how global tectonics and mantle convection are dynamically connected.

The outer shell of Earth comprises an interlocking mosaic of 52 tectonic plates[2]. Among these plates, two groups can be distinguished: a group of seven large plates of similar area covering up to 94% of the planet and a group of smaller plates whose areas follow a fractal distribution[2,3]. The presence of these two statistically distinct groups was previously proposed to reflect two distinct evolutionary laws: the large size group being tied to mantle flow and the other to lithosphere dynamics[3]. In contrast, others studies[5,6] have suggested that this plate layout is produced by superficial processes, because the larger plates may also fit a fractal distribution. Resolving this controversy has been limited by the exclusive use of statistical tools, which do not provide an understanding of the underlying forces and physical principles behind the organization of the plate system.

Here, we use 3D spherical models of mantle convection to uncover the geodynamical processes that drive the tessellation of tectonic plates. Our dynamic models combine pseudo-plasticity and large variations in viscosity (Fig. 1; see Methods), which generate a plate-like behaviour self-consistently[9–11], including fundamental features of sea-floor spreading[12]. In our models, pseudo-plasticity is implemented through a yield stress that represents a plastic limit at which the viscosity drops and strain localization occurs, producing the equivalent of plate boundaries. The value of the yield stress is a measure of the stress at plate

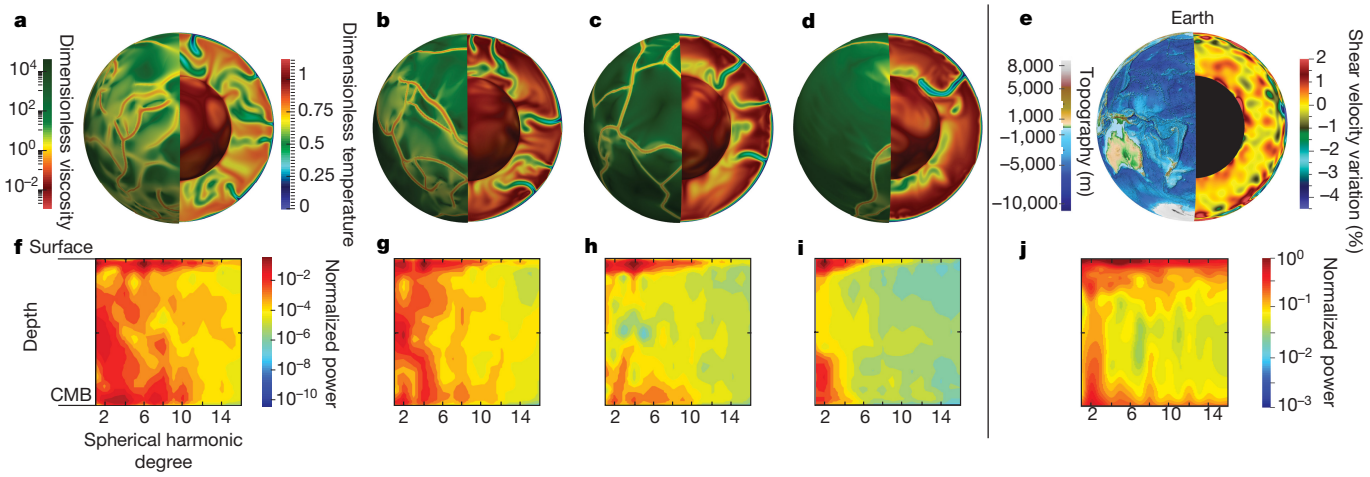


Figure 1 | Snapshots of convection calculations and of Earth with associated spectral heterogeneity maps of the temperature field and seismic velocity field. The spectral heterogeneity maps are normalized by the value of the highest power. **a**, The convection solution with a yield stress of 100 MPa contains a large number of plate boundaries. **f**, The corresponding spherical harmonic map is dominated by degree 6 in the shallow boundary layer. **b**, The convection solution with a yield stress of 150 MPa has fewer plate boundaries and a decreasing number of slabs. **g**, The corresponding spherical harmonic map is dominated by degree 4 at the surface. **c**, The convection solution with a yield stress of 200 MPa has even fewer plate boundaries. **h**, The corresponding spherical harmonic map is dominated by degree 4 at the surface. **d**, The convection solution with a yield stress of 250 MPa has a surface that is barely deformed. **i**, The corresponding spherical harmonic map is blue and dominated by degree 2. **e**, ETOPO1[29] global relief model of Earth and a cross-section through S-wave tomographic model SEMUCB-WM1[30]. **j**, The corresponding spherical harmonic map of the tomographic model is dominated by degrees 4–5 at the surface. CMB, core–mantle boundary.

[1]Laboratoire de Géologie de Lyon, École Normale Supérieure, UMR 5276 CNRS, Université de Lyon 1, 69622 Villeurbanne, France. [2]Institut Universitaire de France, 103 Boulevard Saint Michel, 75005 Paris, France. [3]EarthByte Group, School of Geosciences, Madsen Building F09, University of Sydney, New South Wales 2006, Australia. [4]Institute of Geophysics, Department of Earth Sciences, ETH Zürich, Sonneggstrasse 5, 8092 Zurich, Switzerland.

© 2016 Macmillan Publishers Limited. All rights reserved

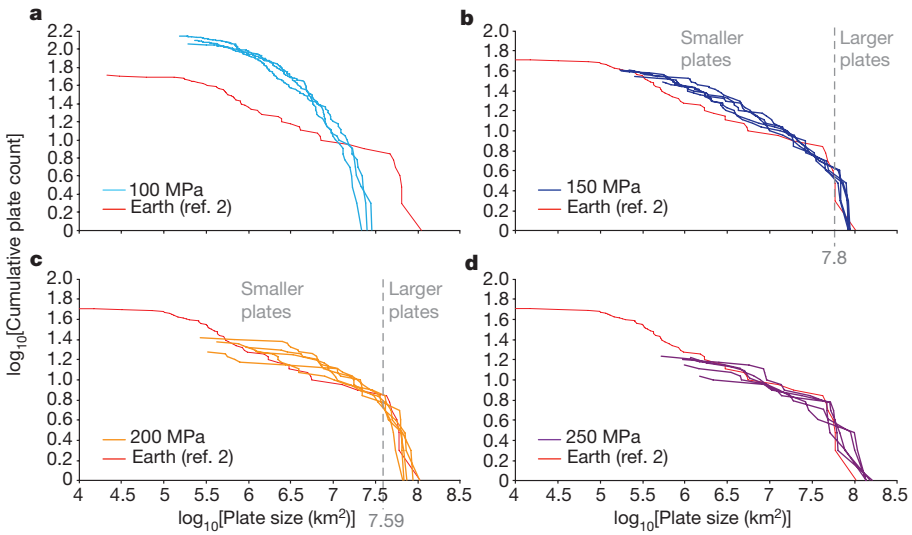


Figure 2 | Plots of the logarithm of cumulative plate count versus the logarithm of plate size for four yield stress values and Earth. The cumulative plate count represents the number of plates that exceed a given area. The graphs contain three data sets for a yield stress of 100 MPa or five data sets for other yield stress values, and the data set for Earth[2], in which the distinction between small plates and large plates (indicated by the vertical dashed lines) is around $10^{7.6}$ km^2 (39,800,000 km^2). **a**, Graph for models with a yield stress of 100 MPa, showing a distribution of small and medium plates. **b**, Graph for models with a yield stress of 150 MPa, showing a distinction between distributions of the large and the small plates. The shift in the distribution occurs at a plate size of about $10^{7.8}$ km^2 (63,100,000 km^2). **c**, Graph for models with a yield stress of 200 MPa, displaying fewer small plates; the groups of small and large plates are distinct and split at about $10^{7.6}$ km^2 (39,800,000 km^2). **d**, Graph for models with a yield stress of 250 MPa, showing only medium and large plates. The division between smaller and larger plates in **b** and **c** corresponds to the cross-over of the fitted slopes of the large and smaller plates (Extended Data Fig. 3).

boundaries and does not necessarily correspond to experimental values. We determine the yield stress range that allows plate-like behaviour, as in previous studies[13–15]. For our convection parameterization, this range exists between 100 MPa, below which surface deformation is very diffuse, and 350 MPa, above which the surface consists of a stagnant lid. We analyse the plate pattern of models with yield stresses of 100 MPa (model 1), 150 MPa (model 2), 200 MPa (model 3) and 250 MPa (model 4) (see Fig. 1). Typically, 90% of the deformation is concentrated in less than 15% of the surface in our models.

Convection modelling generates continuous fields. As a consequence, we have to use plate tectonics rules to delineate the layouts of plates that self-consistently emerge in our dynamical solutions. We digitize plate boundaries on several snapshots for each yield stress value. To be sure that we study snapshots that are substantially different and not correlated, we pick snapshots separated by more than 100 Myr (ref. 16). We choose three snapshots for model 1 and five snapshots for every other model (see Methods). We manually build plate polygons using GPlates[17] through a careful analysis of the surface velocity, horizontal divergence, viscosity, synthetic sea-floor age and temperature field for each snapshot (see Methods, Extended Data Figs 1 and 2). From this we extract the cumulative number versus area distribution of plates for each convection snapshot (Fig. 2).

In model 1 (Fig. 2a), there are more than a hundred plates distributed along a smooth curve. The smallest plate has a size similar to the Easter microplate and the largest one is smaller than the South American plate, which is notably smaller than Earth's larger plates. In contrast, the largest plate for model 4 is larger than the Pacific plate and small plates are absent (Fig. 2d). The snapshots of the models with intermediate yield stresses (models 2 and 3) display the same distributions of plate sizes as observed on Earth (Fig. 2b, c, Extended Data Fig. 3). For a yield stress of 150 MPa (Fig. 2b), the smallest plate is the equivalent of the South Sandwich microplate and the size of the largest is between the areas of the North American plate and the Pacific plate. For a yield stress of 200 MPa (Fig. 2d), the smallest plate is slightly larger than that for a yield stress of 150 MPa, but the largest plate is close in area to the Pacific plate.

Our models indicate that the maximum plate size increases with increasing yield stress, which itself has the effect of increasing the wavelength of convection[15]. For the lowest yield stress value the spherical harmonic power spectrum of the temperature field is dominated by shorter wavelengths and by spherical harmonic degree 6 in the shallow boundary layer (Fig. 1f), which represents the existence of numerous subduction zones and the relatively short wavelengths of the flow in the mantle. For the two intermediate values of 150 MPa and 200 MPa (Fig. 1g, h) the spectra drift to longer wavelengths because degree 4 dominates in the shallow boundary layer, corresponding to a lower number of subduction zones. The maximum size of the plates is similar in both cases. When the yield stress increases to 250 MPa (Fig. 1i), degree 2 dominates in the shallow boundary layer, corresponding to the maximum size of the plates in all of the models. These results suggest that the size of the large plates follows the spacing between active downwellings.

Previous studies on the distributions of smaller plates point to a fragmentation process[5]. We therefore focus on triple junctions, which are symptoms of plate fragmentation: the splitting of a plate into two smaller ones necessarily produces two triple junctions. Both the models and Earth display considerably more triple junctions on subduction zones than on mid-ocean ridges (106.6 versus 75.6 on

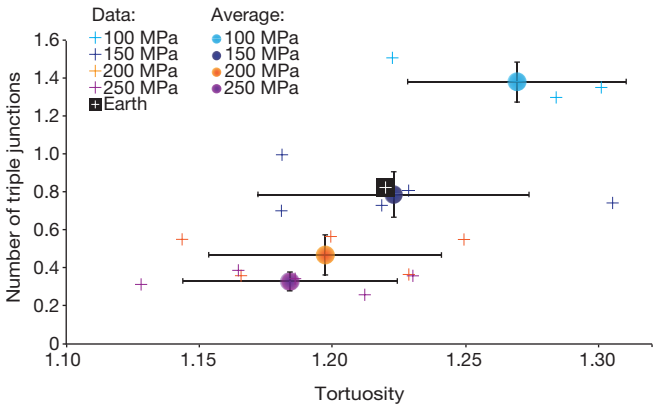


Figure 3 | Number of triple junctions per 1,000 km of subduction zones versus the average tortuosity. Data are shown for four yield stress values and Earth (see legend). The tortuosity is the ratio of the length of the subduction zone to the length of the great circle between the end points. The error bars represent the standard deviation for each data set.

© 2016 Macmillan Publishers Limited. All rights reserved

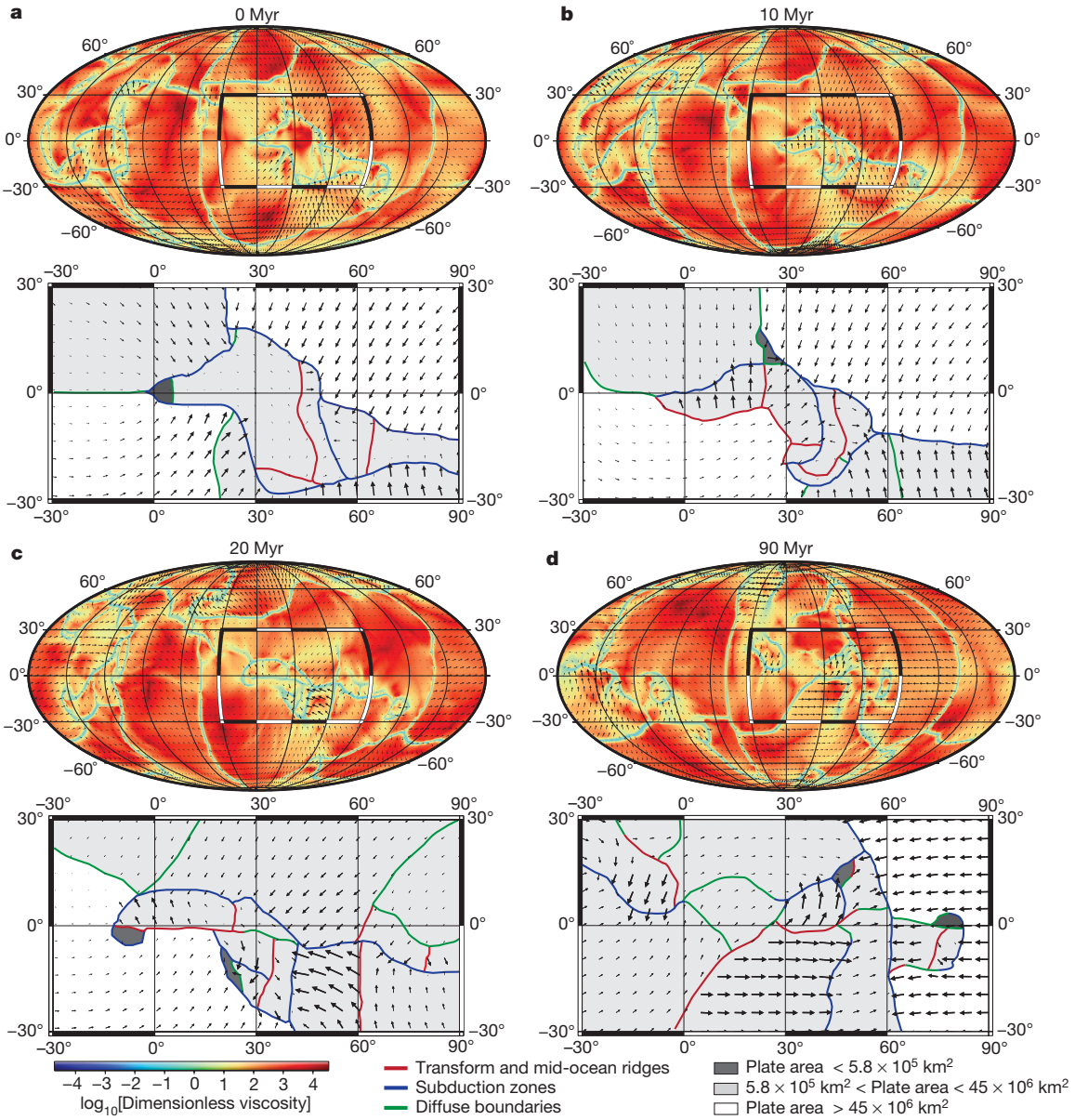


Figure 4 | Global viscosity maps of model 2 and the associated kinematics. a–c, Maps are separated by 10 Myr. The shapes of the large plates do not change much, whereas the adjustment of the small plates evolves quickly. **d**, 90 Myr after the first snapshot (**a**), the distribution of the large plates and smaller plates has evolved substantially. In **a–d**, the top panels show the viscosity of the mantle (colour scale); the bottom panels show the different boundary types (coloured lines) and plate sizes (shading) within the boxed regions in the top panels (which focus on longitudes between −30° and 90° and latitudes between −30° and 30°). The arrows indicate the direction and magnitude (represented by arrow length) of the mantle flow. Plate-size categories are determined in Extended Data Fig. 3.

average for model 2; 131 versus 71 on Earth today), despite the fact that mid-ocean ridges are more elongated than trenches (total length of mid-ocean ridges and trenches: 79,000 km versus 66,000 km on average for model 2; 72,500 km versus 48,000 km on Earth today). Likewise, the triple junctions that are mainly composed of trench segments are those that involve smaller plates in higher proportions (Extended Data Fig. 4). Hence subduction zones focus fragmentation and the formation of smaller plates. On Earth, only the Galapagos, Easter and Juan Fernandez plates formed away from any trench or collisional area.

Our calculations show that plates fragment mostly in connection with curved trenches. Indeed, surface velocities tend to be perpendicular to the trench where slabs sink. Therefore a bend in the trench corresponds to differential motion and hence high stresses. As a consequence, a concave plate under tensile stresses fragments and triple junctions connect the trench with new ridge/transform/diffuse segments. This is consistent with the observed correlation between the tortuosity of trenches and the number of triple junctions per unit length of subduction (Fig. 3). Because increasing the yield stress produces less tortuous trenches and fewer triple junctions per unit length of trench, smaller plate generation is also controlled by the strength of the lithosphere.

The models with plate area distributions similar to Earth also have similar lengths of convergent boundaries to Earth, shown by comparing the trenches in our models with trenches plus mountain belts on Earth[2]. Moreover, the computed temperature heterogeneity spectra of the intermediate yield stress case (Fig. 1g) are consistent with tomographic models of Earth's mantle[18] (Fig. 1j), having degree 2 dominating in the deep mantle. However, our models include simplifications because of computational limitations: a lower Rayleigh number than on Earth (10^6 versus about 10^7), incompressibility and no chemical differences (no continents or deep chemical piles). The physics principles

© 2016 Macmillan Publishers Limited. All rights reserved

we propose for the plate size tessellation are not specifically dependent on the Rayleigh number[19], although the yield stress values could be different. Compressibility should have little impact on the surface tectonics because it concerns the deeper flow[20]. The addition of continents, which help to generate more Earth-like area–age distributions of the sea floor[12], should reinforce the presence of the larger plates and ensure large-scale flow.

On the basis of our results, we propose that the plate pattern on Earth is produced by the dynamic feedback between mantle convection and the strength of the lithosphere. The self-organized subduction structure defines the pattern of large and small plates through slab pull and suction. The large-plate system evolves over hundreds of millions of years through global reorganizations of mantle flow due to the initiation and shutdown of subduction (Fig. 4). This timescale is commensurate with the lifetime of slabs[21]. In contrast, the smaller plates in our models evolve on shorter timescales of tens of millions of years (Fig. 4). They record lateral changes in trench geometry and slab migrations[22]. The enhanced sensitivity of the smaller plates to the readjustment of subduction systems is consistent with present-day observations of seafloor spreading in many back-arc regions. Our models also reveal that global and regional changes in plate motions may be more readily and dramatically expressed in these smaller plates than in the larger plates. For instance, the Parece Vela and Shikoku basins in the Philippine Sea plate record a major clockwise change in the spreading direction between 22 Myr ago and 23 Myr ago (ref. 7), at the same time that the larger Pacific plate records substantial plate boundary and plate motion changes (for example, the fragmentation of the Farallon plate[23] and the collision of the Ontong Java Plateau with the Melanesian subduction zone[24]). In the same way, the Lau Basin in the southwest Pacific initiated its main spreading phase by successive southward propagation around 4 Myr ago (ref. 8), at the same time as a change in the spreading direction in the northeast[25] and southwest Pacific[26] and a major phase of subsidence across the Atlantic[27].

We propose that the plate layout is a property that characterizes a dynamic feedback between mantle convection and lithosphere strength. The larger plates are an expression of the dominating convection wavelength, and their fragmentation into smaller plates is driven by subduction geometry. The decreasing number of smaller plates in pre-Cenozoic-era tectonic reconstructions[3,4] is therefore an artificial consequence of the diminishing quantity of preserved sea floor. Confirming the existence of migrating intra-oceanic subduction systems such as in Panthalassa[28] may help to correct that bias. Over longer geologic timescales, the size distribution of plates has certainly evolved with the slow cooling of Earth. Following the declining convective vigour, the lithosphere gets stronger relative to mantle forces. Therefore, this study suggests that since plate tectonics started on Earth, it may have operated with fewer, larger plates as the planet has cooled down.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 September 2015; accepted 4 April 2016.
Published online 15 June 2016.

1. Le Pichon, X. Sea-floor spreading and continental drift. *J. Geophys. Res.* **73,** 3661–3697 (1968).
2. Bird, P. An updated digital model of plate boundaries. *Geochem. Geophys. Geosyst.* **4,** 1027 (2003).
3. Morra, G., Seton, M., Quevedo, L. & Müller, R. D. Organization of the tectonic plates in the last 200 Myr. *Earth Planet. Sci. Lett.* **373,** 93–101 (2013).
4. Seton, M. *et al.* Global continental and ocean basin reconstructions since 200 Ma. *Earth Sci. Rev.* **113,** 212–270 (2012).
5. Sornette, D. & Pisarenko, V. Fractal plate tectonics. *Geophys. Res. Lett.* **30,** 1105 (2003).
6. Vallianatos, F. & Sammonds, P. Is plate tectonics a case of non-extensive thermodynamics? *Physica A* **389,** 4989–4993 (2010).
7. Sdrolias, M., Roest, W. R. & Müller, R. D. An expression of Philippine Sea plate rotation: the Parece Vela and Shikoku Basins. *Tectonophysics* **394,** 69–86 (2004).
8. Taylor, B., Zellmer, K., Martinez, F. & Goodliffe, A. Sea-floor spreading in the Lau back-arc basin. *Earth Planet. Sci. Lett.* **144,** 35–40 (1996).
9. Moresi, L. & Solomatov, V. Mantle convection with a brittle lithosphere: thoughts on the global tectonic styles of the Earth and Venus. *Geophys. J. Int.* **133,** 669–682 (1998).
10. Trompert, R. & Hansen, U. Mantle convection simulations with rheologies that generate plate-like behaviour. *Nature* **395,** 686–689 (1998).
11. Tackley, P. J. Self-consistent generation of tectonic plates in time-dependent, three-dimensional mantle convection simulations: 1. Pseudoplastic yielding. *Geochem.Geophys. Geosyst.* **1,** 1021 (2000).
12. Coltice, N., Seton, M., Rolf, T., Müller, R. & Tackley, P. J. Convergence of tectonic reconstructions and mantle convection models for significant fluctuations in seafloor spreading. *Earth Planet. Sci. Lett.* **383,** 92–100 (2013).
13. Ricard, Y., Bercovici, D. & Schubert, G. A two-phase model for compaction and damage: 2. Applications to compaction, deformation, and the role of interfacial surface tension. *J. Geophys. Res.* **106,** 8907 (2001).
14. Stein, C., Schmalzl, J. & Hansen, U. The effect of rheological parameters on plate behaviour in a self-consistent model of mantle convection. *Phys. Earth Planet. Inter.* **142,** 225–255 (2004).
15. van Heck, H. J. & Tackley, P. J. Planforms of self-consistently generated plates in 3D spherical geometry. *Geophys. Res. Lett.* **35,** L19312 (2008).
16. Bello, L., Coltice, N., Rolf, T. & Tackley, P. J. On the predictability limit of convection models of the Earth's mantle. *Geochem. Geophys. Geosyst.* **15,** 2319–2328 (2014).
17. Williams, S. E., Müller, R. D. & Landgrebe, T. C. W. An open-source software environment for visualizing and refining plate tectonic reconstructions using high-resolution geological and geophysical data sets. *GSA Today* **22(4),** 4–9 (2012).
18. Becker, T. W. & Boschi, L. A comparison of tomographic and geodynamic mantle models. *Geochem. Geophys. Geosyst.* **3,** 1003 (2002).
19. van Heck, H. J. & Tackley, P. J. Plate tectonics on super-Earths: equally or more likely than on Earth. *Earth Planet. Sci. Lett.* **310,** 252–261 (2011).
20. Tackley, P. J. Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid. *Phys. Earth Planet. Inter.* **171,** 7–18 (2008).
21. Matthews, K. J., Seton, M. & Müller, R. D. A global-scale plate reorganization event at 105–100 Ma. *Earth Planet. Sci. Lett.* **355–356,** 283–298 (2012).
22. Stegman, D. R., Schellart, W. P. & Freeman, J. Competing influences of plate width and far-field boundary conditions on trench migration and morphology of subducted slabs in the upper mantle. *Tectonophysics* **483,** 46–57 (2010).
23. Barckhausen, U., Ranero, C. R., Cande, S. C., Engels, M. & Weinrebe, W. Birth of an intraoceanic spreading center. *Geology* **36,** 767–770 (2008).
24. Petterson, M. G. *et al.* Geological–tectonic framework of Solomon Islands, SW Pacific: crustal accretion and growth within an intra-oceanic setting. *Tectonophysics* **301,** 35–60 (1999).
25. Harbert, W. Late Neogene relative motions of the Pacific and North America plates. *Tectonics* **10,** 1–15 (1991).
26. Tebbens, S. F. & Cande, S. C. Southeast Pacific tectonic evolution from early Oligocene to present. *J. Geophys. Res.* **102(B6),** 12061–12084 (1997).
27. Cloetingh, S. A. P. L., Gradstein, F. M., Kooi, H., Grant, A. C. & Kaminski, M. M. Plate reorganization: a cause of rapid late Neogene subsidence and sedimentation around the North Atlantic? *J. Geol. Soc. Lond.* **147,** 495–506 (1990).
28. van der Meer, D. G., Torsvik, T. H., Spakman, W., van Hinsbergen, D. J. J. & Amaru, M. L. Intra-Panthalassa Ocean subduction zones revealed by fossil arcs and mantle structure. *Nat. Geosci.* **5,** 215–219 (2012).
29. Amante, C. & Eakins, B. W. *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis* www.ngdc.noaa.gov/mgg/global/relief/ETOPO1/image/color_etopo1_ice_low.tif.zip (US National Oceanic and Atmospheric Administration, 2009).
30. French, S. W. & Romanowicz, B. A. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots. *Nature* **525,** 95–99 (2015).

Acknowledgements The research leading to these results was funded by the European Research Council within the framework of the SP2-Ideas Programme ERC-2013-CoG under ERC grant agreement 617588. We thank S. Durant and E. Debayle for helping to make Fig. 1e, i and E. J. Garnero for his inputs. Calculations were performed on the AUGURY supercomputer at P2CHPD Lyon. N.C. was supported by the Institut Universitaire de France. R.D.M and M.S are supported by ARC grants DP130101946 and FT130101564.

Author Contributions C.M. developed the methodology for analysing the convection models, conducted the plate analysis, contributed to the interpretation and wrote the manuscript. N.C. conducted the convection calculations, contributed to the development of the methodology and analysis, contributed to the interpretation and wrote the manuscript. M.S. and R.D.M. provided guidance with GPlates and scripts, contributed to the interpretation and wrote the manuscript. P.J.T. provided the StagYY convection code, guidance on using it and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.M. (claire.mallard@univ-lyon1.fr).

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Convection models. The models computed here have similar parameterizations to those published in ref. 16, except no surface velocities are imposed here (free convection). We solve the non-dimensional equations of mass, momentum and heat conservation in a 3D spherical geometry using the code StagYY (ref. 20). The flow is incompressible under the Boussinesq approximation. Viscosity is the only variable material property in our models. Variations of other material properties (expansion coefficient, thermal diffusivity, heat production) are neglected.

The Rayleigh number Ra is defined here as

$$\mathrm{Ra} = \frac{\rho g \alpha \Delta T L^3}{\kappa \eta_0}$$

where ρ is density, g is the gravitational acceleration, α is the thermal expansivity, ΔT is the temperature drop across mantle depth, L is the mantle thickness, κ is thermal diffusivity and η_0 is the reference viscosity at the base of the mantle. The non-dimensional temperature is set to $T=0$ at the surface and $T=1$ at the base of the mantle. A non-dimensional internal heat production of 20 is chosen, such that the basal heat flux is about 14% of the total. This is in the lower range of the estimates for heat flow at the core–mantle boundary[31].

In our models Ra is 10^6, which is about 10–50 times lower than is expected for Earth, and produces a top boundary layer that is 300-km thick. We were limited to this value of Ra because of the computational power required to solve for convection with large viscosity variations. The average resolution is 45 km laterally and vertically for all of the models.

The viscosity in our models depends on temperature and depth as

$$\eta(T,z) = \eta_z(z)\exp(0.064 + 30/(T+1))$$

where z is the depth. A value of 30 for the non-dimensional activation energy produces six orders of magnitude variations in viscosity with temperature.

The depth dependence of viscosity is taken into account such that

$$\eta_z(z) = a\exp\left\{\ln(B)\left[1 - 0.5\left(1 - \tanh\left(\frac{d_0 - z}{d_{\mathrm{step}}}\right)\right)\right]\right\}$$

where B is the factor of the viscosity jump at depth d_0 over a thickness $2d_{\mathrm{step}}$, and a is a prefactor ensuring $\eta_z = \eta_0$ for temperature $T=1$ at the base of the mantle. Based on geoid[32] and post-glacial rebound modelling[33], B is set to 30 here and the viscosity jump occurs between 750-km and 850-km depth (d_0 is 0.276 and d_{step} is 0.02).

Pseudo-plasticity is implemented through a stress dependence of the viscosity with yield stress[9–11]. When the local stress reaches the yield stress value σ_y, the viscosity is computed as

$$\eta = \frac{\sigma_y}{2\dot{\epsilon}}$$

where $\dot{\epsilon}$ is the second invariant of the strain-rate tensor. The StagYY code has been benchmarked with such rheology[34]. Yield stress is the only parameter varied in this study. Taking $\eta_0 = 10^{23}$ Pa s, the yield stress values that produce plate-like behaviour are between 100 MPa and 350 MPa.

In our models, the viscosity drops by a factor of 10 in the vicinity of ridges, where the temperature crosses the solidus temperature, given by a simple linear model $T_{\mathrm{sol}} = 0.6 + 7.5z$, and without a dependence on the melt fraction. This effect improves slightly plate-like behaviour and has been used in previous studies[11,15]. The models are started from *ad hoc* initial conditions, and run for up to five billion years to ensure a statistical steady state and the stability of the dynamic regime. Such long runs ensure that the initial conditions are forgotten so they don't affect the outcome results. From the solutions at a statistical steady state, we compute the dynamic evolutions of the models that are analysed in this study.

Code availability. The code StagYY is the property of P.J.T. and Eidgenössische Technische Hochschule (ETH) Zürich and is available on request from P.J.T. (paul.tackley@erdw.ethz.ch).

Building tectonic plates. We established a method to define the boundaries and the geometry of tectonic plates on the surface of our convection models. First the boundaries need to be identified to define the outline of the plates themselves (plate polygons). The same method was applied for each of the 18 snapshots of the models we present. This is a relatively small sample because the precise determination of the plate layout for one snapshot is very time-consuming. Only three snapshots have been studied for model 1 because of the large number of plates (more than 100). The GPlates software is used to trace all plate boundaries, interactively building digital plate-tectonic layouts.

Identification of major boundaries. The first step here is to identify the major and localized boundaries on the surface of the convection models. We use the viscosity, temperature and velocity data. The maps of the sea-floor ages obtained from the heat flux (Extended Data Fig. 1a) allow the youngest zones (0 Myr old) to be identified as mid-ocean ridges and the oldest zones (180–280 Myr old) as subduction zones. In the same manner, we use maps of the horizontal divergence (Extended Data Fig. 1b) inferred from the surface velocities. Hence, the divergence zones show the localization of the mid-ocean ridges for dimensionless divergence values between 0 and 30,000 and the convergence zones show the subduction zones with values between −15,000 and 0. Transform zones (as our model is continuous, there are no faults but shear zones) exist in our models and are identified via surface vorticity maps. To minimize the time it takes to interactively build plate boundary models, mid-ocean ridges and transform zones are included in the same group of boundaries. Nevertheless, for the model with a yield stress of 150 MPa, we computed a length of about 79,000 km for mid-ocean ridges on average and a length of about 2,600 km for transform regions. In comparison, these lengths on Earth are 67,000 km for mid-ocean ridges and 5,131 km for transform regions.

The identification of these two types of major boundaries (subduction zones and mid-ocean ridges) does not always allow us to close polygons to obtain tectonic plates. Even if some boundaries can be extrapolated, many zones necessitate more thorough work, as discussed below.

Identification of diffuse boundaries. To close polygons, other boundaries need to be defined. The study of deviatoric stress allows us to identify some diffuse junctions. In the models, non-yielded boundaries are set between two zones where there is little change in the velocity vector. They exist in ductile zones that are visible as a result of a fan of velocity vectors (Extended Data Fig. 2). This geometric configuration implies a large zone of deformation similar to that found for intraplate deformation, which is defined as a diffuse boundary. That is exactly the definition of diffuse boundaries on Earth[35]. The delimitation of the diffuse boundaries between two zones with different velocities implies a non-negligible error in the estimation of the Euler pole (and the calculated velocities) that we quantify.

The identification of these three types of boundaries (mid-ocean ridges, subduction zones and diffuse boundaries) allows us to close topological polygons defined by these boundaries (Extended Data Fig. 1c). These polygons are tectonic plates, but before they can be used we need to evaluate the error we made in the delimitation of tectonic plates according to the plate tectonic theory.

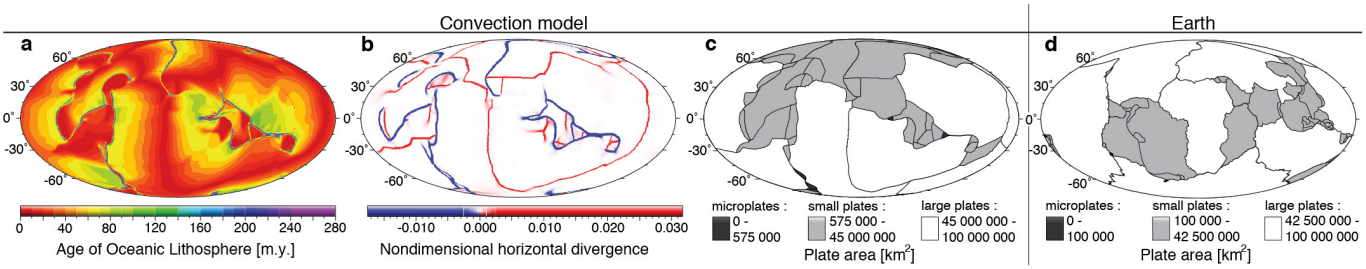
Fit of the plate model with the convection model. We compare the raw velocity data of the convection models with the *a posteriori* velocities calculated using Euler's theorem for the corresponding plate layout. We first extract the raw velocity data for each plate using the plate polygons determined previously. We then use the raw velocities to invert the angular velocity vector using the inverse method described in ref. 36, and compute the predicted velocities on the basis of the inverted angular velocity vector. As a measure of the quality of the fit of our plate model to the convection model, we compute the plateness P of the plate layout following ref. 37

$$P = 1 - \Delta V_{\mathrm{rms}} / V_{\mathrm{rms}}$$

where ΔV_{rms} is the root-mean-square difference between the velocities of the convection model and those predicted with plate rotations, and V_{rms} is the root-mean-square surface velocity of the model. We obtain values of P between 0.75 and 0.81 (1 would be perfectly rigid plates, 0 would absolutely preclude the use of plate approximation), which is consistent with the fact that 90% of the deformation is concentrated in 15% of the surface of the models.

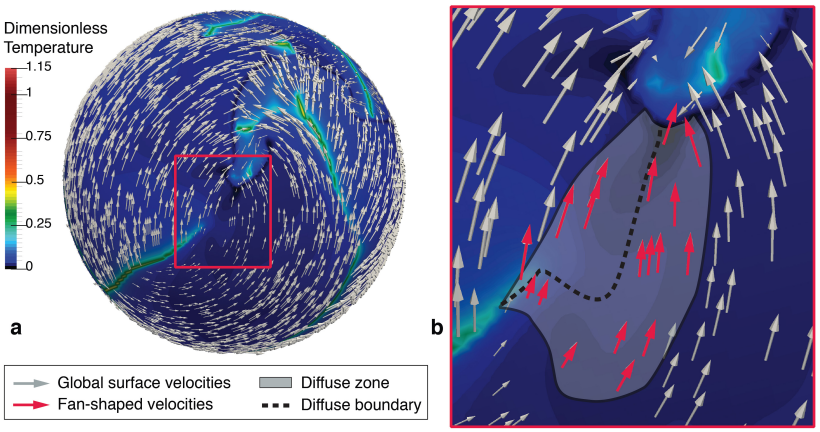
31. Lay, T., Hernlund, J. & Buffett, B. A. Core–mantle boundary heat flow. *Nat. Geosci.* **1,** 25–32 (2008).
32. Ricard, Y., Richards, M., Lithgow-Bertelloni, C. & Le Stunff, Y. A geodynamic model of mantle density heterogeneity. *J. Geophys. Res.* **98,** 21895–21909 (1993).
33. Mitrovica, J. X. Haskell [1935] revisited. *J. Geophys. Res.* **101,** 555–569 (1996).
34. Tosi, N. *et al.* A community benchmark for viscoplastic thermal convection in a 2-D square box. *Geochem. Geophys. Geosyst.* **16,** 2175–2196 (2015).
35. Gordon, R. G. in *The History and Dynamics of Global Plate Motions* (eds Richards, M. A. *et al.*) 143–159 (Geophysical Monograph Series, Vol. 121, American Geophysical Union, 2000).
36. Goudarzi, M. A., Cocard, M. & Santerre, R. EPC: Matlab software to estimate Euler pole parameters. *GPS Solut.* **18,** 153–162 (2014).
37. Zhong, S., Gurnis, M. & Moresi, L. Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models. *J. Geophys. Res.* **103,** 15255–15268 (1998).

© 2016 Macmillan Publishers Limited. All rights reserved



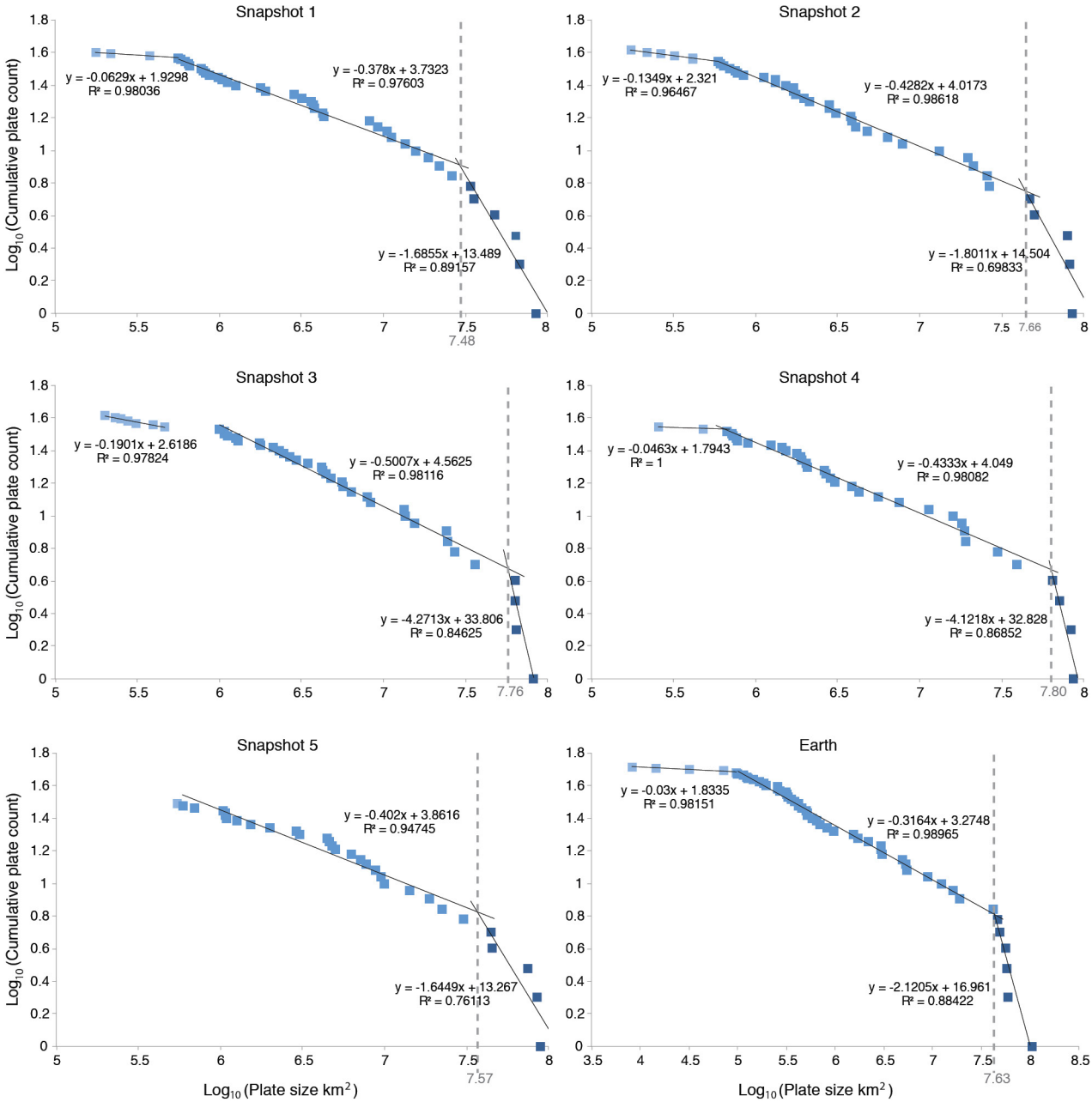
Extended Data Figure 1 | Maps of the surface of a snapshot from a convection model with a yield stress of 150 MPa and of the plate layout of Earth. a, Map of sea-floor age with the youngest ages in red characteristic of mid-ocean ridges and the oldest zones in blue characteristic of subduction zones. m.y., millions of years. **b**, Map of non-dimensional horizontal divergence, with divergence zones (mid-ocean ridges) shown in red and convergence zones (subduction zones) in blue. **c**, **d**, Maps of the plate sizes of the convection model (**c**) and Earth (**d**). The plate size categories are determined in Extended Data Fig. 3.

© 2016 Macmillan Publishers Limited. All rights reserved



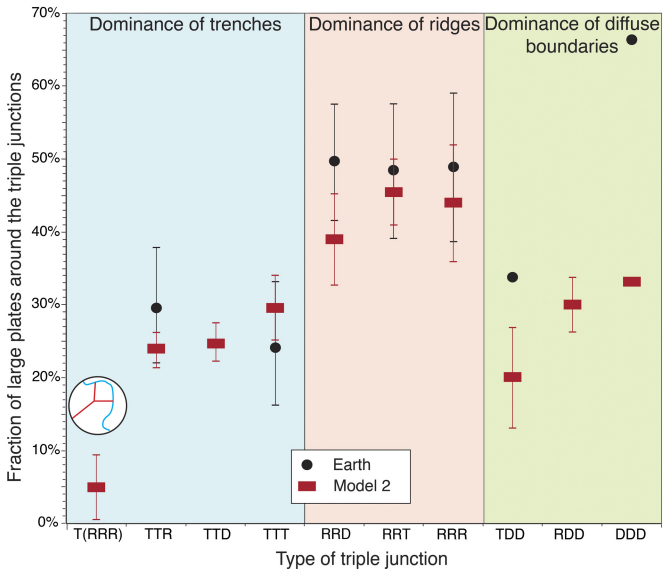
Extended Data Figure 2 | Subsurface temperature of a convection model with a yield stress of 150 MPa showing a diffuse plate boundary. **a**, Global temperature (colour scale) and surface velocities (arrows). The dark zones represent subduction zones and the light zones indicate mid-ocean ridges. **b**, Zoom-in of the red boxed region in **a** showing a diffuse boundary; the steady lateral change of velocity directions (red arrows) characterizes the intraplate diffuse zone (grey shaded area), allowing the determination of a diffuse boundary (black dashed line).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 3 | Plots of the logarithm of the cumulative plate count versus the logarithm of the plate size for the snapshots of model 2 and Earth. The data for Earth is taken from ref. 2. The plots show the distribution of microplates in light blue, small plates in mid-blue and large plates in dark blue. The equations of the black fit lines and the correlation coefficients R^2 are also shown.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Plot of the fraction of large plates adjoining a triple junction versus the type of triple junction for model 2 and for Earth. The data for Earth is taken from ref. 2. The red rectangles correspond to model 2 and the black circles to Earth. The coloured backgrounds indicate of dominance of each boundary type: blue shows triple junctions that are mainly composed of subduction zones, red shows the dominance of mid-ocean ridges or transform boundaries and green the dominance of diffuse boundaries. T, trenches; R, ridges; D, diffuse boundary. We added a type of triple junction T(RRR); these triple junctions are directly connected to curved trenches and produce back-arc basins with small plates, hence they are included in the area of the plot dominated by subduction zones. The error bars represent the standard deviation of the fraction of large plates around a triple junction for model 2 and Earth.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18326

Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation

Jos Barlow[1,2,3], Gareth D. Lennox[1], Joice Ferreira[4], Erika Berenguer[1], Alexander C. Lees[2,5], Ralph Mac Nally[6], James R. Thomson[6,7], Silvio Frosini de Barros Ferraz[8], Julio Louzada[1,3], Victor Hugo Fonseca Oliveira[1,3], Luke Parry[1,9], Ricardo Ribeiro de Castro Solar[10], Ima C. G. Vieira[2], Luiz E. O. C. Aragão[11,12], Rodrigo Anzolin Begotti[8], Rodrigo F. Braga[3], Thiago Moreira Cardoso[4], Raimundo Cosme de Oliveira Jr[4], Carlos M. Souza Jr[13], Nárgila G. Moura[2,5], Sâmia Serra Nunes[13], João Victor Siqueira[13], Renata Pardini[14], Juliana M. Silveira[1,3], Fernando Z. Vaz-de-Mello[15], Ruan Carlo Stulpen Veiga[16], Adriano Venturieri[4] & Toby A. Gardner[17,18]

Concerted political attention has focused on reducing deforestation[1–3], and this remains the cornerstone of most biodiversity conservation strategies[4–6]. However, maintaining forest cover may not reduce anthropogenic forest disturbances, which are rarely considered in conservation programmes[6]. These disturbances occur both within forests, including selective logging and wildfires[7,8], and at the landscape level, through edge, area and isolation effects[9]. Until now, the combined effect of anthropogenic disturbance on the conservation value of remnant primary forests has remained unknown, making it impossible to assess the relative importance of forest disturbance and forest loss. Here we address these knowledge gaps using a large data set of plants, birds and dung beetles (1,538, 460 and 156 species, respectively) sampled in 36 catchments in the Brazilian state of Pará. Catchments retaining more than 69–80% forest cover lost more conservation value from disturbance than from forest loss. For example, a 20% loss of primary forest, the maximum level of deforestation allowed on Amazonian properties under Brazil's Forest Code[5], resulted in a 39–54% loss of conservation value: 96–171% more than expected without considering disturbance effects. We extrapolated the disturbance-mediated loss of conservation value throughout Pará, which covers 25% of the Brazilian Amazon. Although disturbed forests retained considerable conservation value compared with deforested areas, the toll of disturbance outside Pará's strictly protected areas is equivalent to the loss of 92,000–139,000 km^2 of primary forest. Even this lowest estimate is greater than the area deforested across the entire Brazilian Amazon between 2006 and 2015 (ref. 10). Species distribution models showed that both landscape and within-forest disturbances contributed to biodiversity loss, with the greatest negative effects on species of high conservation and functional value. These results demonstrate an urgent need for policy interventions that go beyond the maintenance of forest cover to safeguard the hyper-diversity of tropical forest ecosystems.

Protecting tropical forests is a fundamental pillar of many national and international strategies for conserving biodiversity[4–6]. Although improved regulatory and incentive measures have reduced deforestation rates in some tropical nations[1,11,12], the conservation value of the world's remaining primary forests may be undermined by the additional impacts of disturbance, which falls into two broad categories (see Methods). First, landscape disturbance results from deforestation itself, with area, isolation and edge effects degrading the condition of the remaining forests[9]. Second, within-forest disturbance, such as wildfires and selective logging, induces marked changes in forest structure and species composition[8,13].

Although the biodiversity consequences of both forms of disturbance are well studied, previous research has overwhelmingly focused on identifying the isolated effects of specific types of disturbance[14,15]. Such studies provide an incomplete understanding of the total disturbance-mediated loss of conservation value arising from multiple interacting drivers[16] and are unable to quantify the extent to which reducing forest loss will succeed in protecting tropical forest biodiversity. Addressing these knowledge gaps is vital for informing forest management strategies in tropical nations, not least because within-forest disturbance can increase even as deforestation rates fall[7,12,17] and thus requires different policy interventions (Extended Data Table 1).

We estimated the combined effects of landscape and within-forest disturbance on biodiversity in primary forests and compared these impacts to the biodiversity loss expected in deforested areas, offering, to our knowledge, the first such analysis for anywhere in the world. Our study focused on two large (>10,000 km^2) frontier regions of the Brazilian Amazon: Paragominas and Santarém, located in the state of Pará (see Methods). Large- and small-stemmed plants, birds and dung beetles were sampled in 371 plots in 36 study catchments distributed along a landscape deforestation gradient (0–94%) (Extended Data Fig. 1). A total of 31 catchments contained remnant primary forests. Within these catchments, we sampled 175 primary forest plots. Of these, 145 had visible evidence of within-forest disturbance (logging and/or fire). The remaining 30 had no evidence of within-forest disturbance and, being located in the largest remaining forest blocks, had minimal landscape disturbance[18,19] (see Methods). Irrespective of their disturbance history, these primary forest plots held considerably more forest species than all other major land-uses (Extended Data Fig. 2).

We used the sum of forest species presences in primary forest plots to estimate the conservation value of a catchment (see Methods). As plots

[1]Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK. [2]MCTI/Museu Paraense Emílio Goeldi, CP 399, Belém, Pará, CEP 66040-170, Brazil. [3]Universidade Federal de Lavras, Setor de Ecologia e Conservação. Lavras, Minas Gerais, CEP 37200-000, Brazil. [4]EMBRAPA Amazônia Oriental. Belém, Pará, CEP 66095-100, Brazil. [5]Cornell Lab of Ornithology, Cornell University, Ithaca, New York 14850, USA. [6]Institute for Applied Ecology, University of Canberra, Bruce, Australian Capital Territory 2617, Australia. [7]Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, 123 Brown Street, Heidelberg, Victoria 3084, Australia. [8]Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Esalq/USP, Avenida Pádua Dias, 11, São Dimas, Piracicaba, SP, CEP 13418-900, Brazil. [9]Universidade Federal do Pará (UFPA), Núcleo de Altos Estudos Amazonicos (NAEA), Av. Perimetral, Numero 1, Guamá, Belém-Pará, CEP 66075-750, Brazil. [10]Universidade Federal de Viçosa, Departamento de Biologia Geral. Av. PH Rolfs s/n. Viçosa, Minas Gerais, CEP 36570-900, Brazil. [11]Tropical Ecosystems and Environmental Sciences Group (TREES), Remote Sensing Division, National Institute for Space Research (INPE), Avenida dos Astronautas, 1.758, Jd. Granja, São José dos Campos, CEP 12227-010, SP, Brazil. [12]College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK. [13]IMAZON, Rua Dom Romualdo de Seixas 1698, Edifício Zion, 11 andar, CEP 66055-200 Belém, PA, Brazil. [14]Instituto de Biociencias, Universidade de São Paulo, Rua do Matão, Travessa 14, 101, CEP 05508-090 São Paulo, Brazil. [15]Universidade Federal de Mato Grosso, Instituto de Biociencias, Departamento de Biologia e Zoologia. Av. Fernando Correa da Costa, 2367, Boa Esperança, CEP 78060-900, Cuiaba, MT, Brazil. [16]Instituto Socio Ambiental Serra do Mar (ISASM), Estrada Ribeirão das Voltas s/n, Lumiar, CEP 28616-010, Nova Friburgo, Brazil. [17]Stockholm Environment Institute, Linnégatan 87D, Box 24218, Stockholm 104 51, Sweden. [18]International Institute for Sustainability, Estrada Dona Castorina, 124, Horto, Rio de Janeiro, 22460-320, Brazil.

© 2016 Macmillan Publishers Limited. All rights reserved

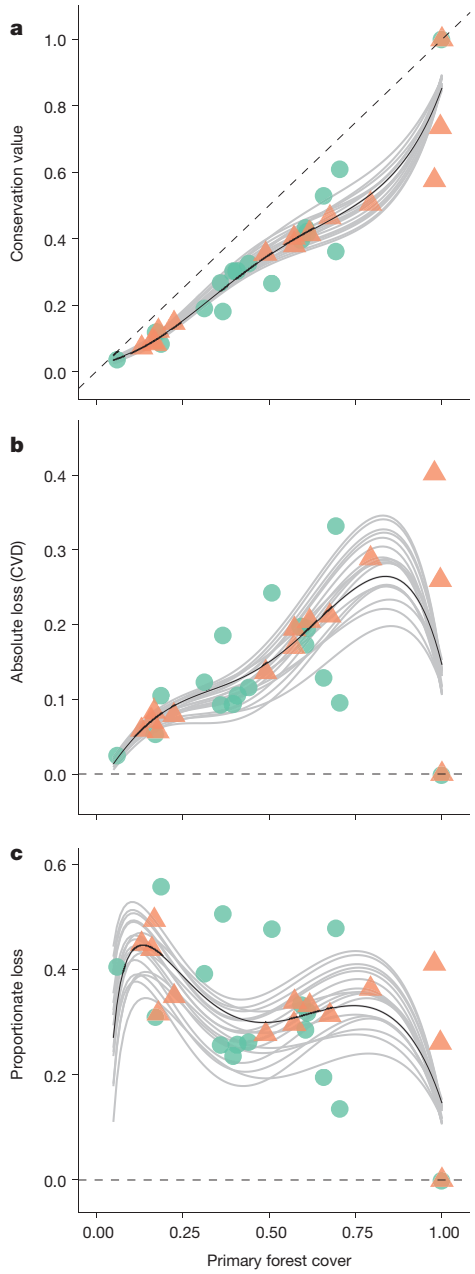


Figure 1 | The conservation status of primary forests. a, Conservation value in Paragominas (circles) and Santarém (triangles). **b**, Total loss of conservation value due to disturbance. **c**, Total loss of conservation value due to disturbance expressed as a proportion of the expected conservation value without disturbance. Dashed lines show expectations without disturbance. Grey lines show all regressions, with the black solid line showing the median response (see Methods). Values were standardized across study regions. There was no significant difference in conservation values between regions in the median response ($F_{1;26} = 1.45$, $P = 0.24$, analysis of covariance (ANCOVA)).

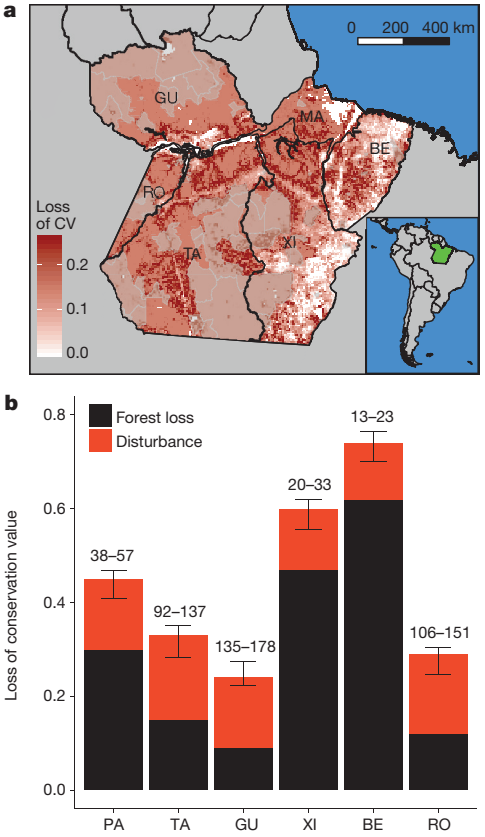


Figure 2 | Conservation value deficit over large spatial scales. **a**, Proportionate loss of conservation value (CV) from disturbance in Pará (median estimate; see Methods). Areas of endemism (AoE) are: Belém (BE), Guiana (GU), Rondônia (RO), Tapajós (TA) and Xingu (XI). These do not include the island of Marajó (MA). Grey shading denotes strictly protected areas. **b**, Proportionate loss of CV in Pará (PA) and its AoEs from forest loss and disturbance (median estimate). Error bars show the range over all approaches to estimating conservation value (see Methods). Numbers show disturbance relative to forest loss (percentage range over approaches).

were allocated in proportion to catchment forest cover, this measure is equivalent to the mean species richness (per unit area) in primary forests multiplied by the proportion of primary forest cover. In the absence of landscape or within-forest disturbances, the expectation of conservation value should respond linearly to forest cover, with slope equal to mean species density (see Methods). The difference between this linear expectation and the observed conservation value of the remaining primary forest provides an estimate of the total biodiversity impact of all landscape and within-forest disturbance. We refer to this difference as the conservation value deficit (CVD). We take a variety of approaches to calculating the CVD, reflecting different ways of classifying forest species, weighting their conservation value and calculating species density in undisturbed forest (see Methods). Here, we report median results from our sensitivity analysis along with the lower and upper bound range. Full results are shown in Fig. 1 and Extended Data Figs 3 and 4.

The conservation value of the remaining primary forests was lower than expected along the entire deforestation gradient. The CVD was unimodal with forest cover, reaching its maximum in catchments with 83% of their primary forests. These catchments retained just 58% of their conservation value (range: 48 to 65%) (Fig. 1a). The CVD was relatively small at low levels of forest cover (Fig. 1b). However, disturbance caused the greatest proportionate loss of conservation value in these catchments, accounting for an approximately 20–50% shortfall in the level of biodiversity that would be predicted for undisturbed forests (Fig. 1c). The robustness of our estimates of the CVD was supported by the similarity of responses across study regions (Fig. 1) and sampled taxa (Extended Data Fig. 3).

The relationship we derived between forest cover and conservation value allowed us, for the first time, to estimate the additional total effect of forest disturbance over large spatial scales. We therefore mapped the disturbance-induced loss of conservation value (CVD) across Pará, which covers $1.26 \times 10^6\,km^2$. We divided the state into grid cells approximately equal in area to our study catchments ($\sim 50\,km^2$). In total, 73% of the ~26,000 cells covering the state were located in private lands or sustainable-use reserves. For these locations, which are most

© 2016 Macmillan Publishers Limited. All rights reserved

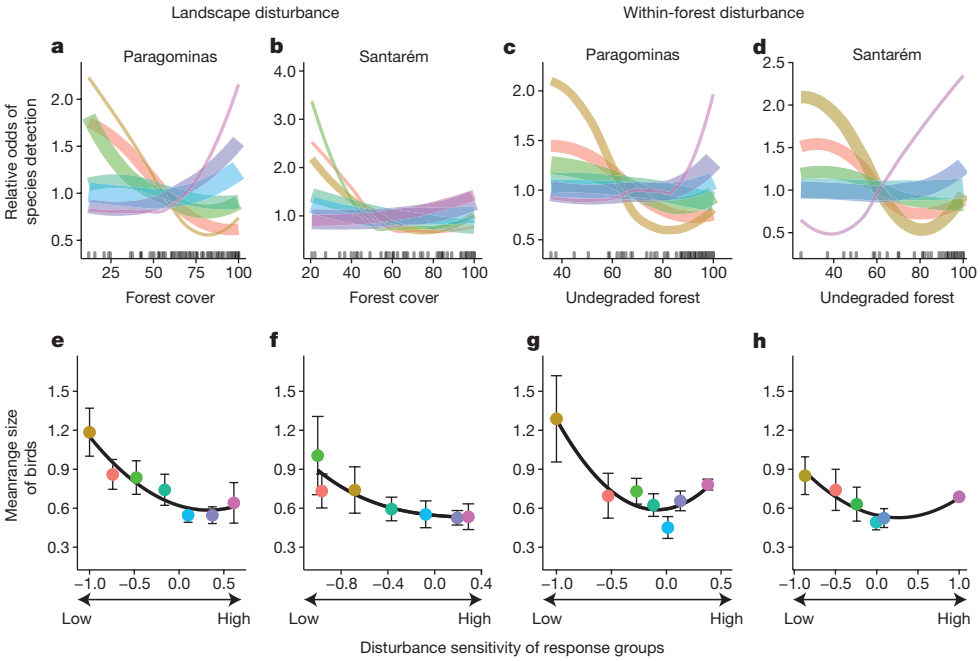


Figure 3 | Response of forest birds to disturbance. a–d, The odds of detecting species groups along gradients of landscape (**a**, **b**) and within-forest (**c**, **d**) disturbance in Paragominas (**a**, **c**) and Santarém (**b**, **d**) (see Methods). Species groups, shown by different coloured lines, are composed of species with similar disturbance responses (see Methods). Line thickness represents the relative size of the groups. **e–h**, Disturbance sensitivity of the species groups related to their mean range size (10^7 km^2). Error bars shows s.e.m. Group colours correspond to groupings in **a–d**. Black lines show significant relationships ($P < 0.05$, F-test) (see Methods).

comparable to our study catchments, the total CVD was equivalent to ~123,000 km^2 of forest loss (range: 92,000 to 139,000 km^2). To put this figure in context, it is 51% (range: 38 to 57%) of the total area deforested across Pará to date (Extended Data Table 2).

Our state-wide analysis revealed considerable spatial variation in the CVD, reflecting differences in deforestation histories (Fig. 2a). We illustrate this variation by estimating the additional loss of conservation value due to disturbance across Pará's five major biogeographic zones (areas of endemism[20], AoE). Median disturbance impacts outweighed biodiversity losses in deforested areas alone in three of the five AoEs (Fig. 2b). The high relative impact of disturbance is shown in the Guiana AoE, where the predicted loss of conservation value from disturbance was 135–178% of the losses estimated in deforested areas. The relative impact of disturbance was lowest in the Belém AoE, which has lost 62% of its native forest cover and is the most deforested AoE in Amazonia. Nonetheless, overall disturbance effects reduced Belém's estimated conservation value from 38% when based on forest cover alone to just 26% (range: 24 to 30%).

The widespread and substantial depletion of conservation value in remaining primary forests highlights the pressing need for policies that target the most prominent drivers of disturbance-induced biodiversity loss. Although measures to combat deforestation may help limit landscape disturbance, they rarely consider the spatial configuration of

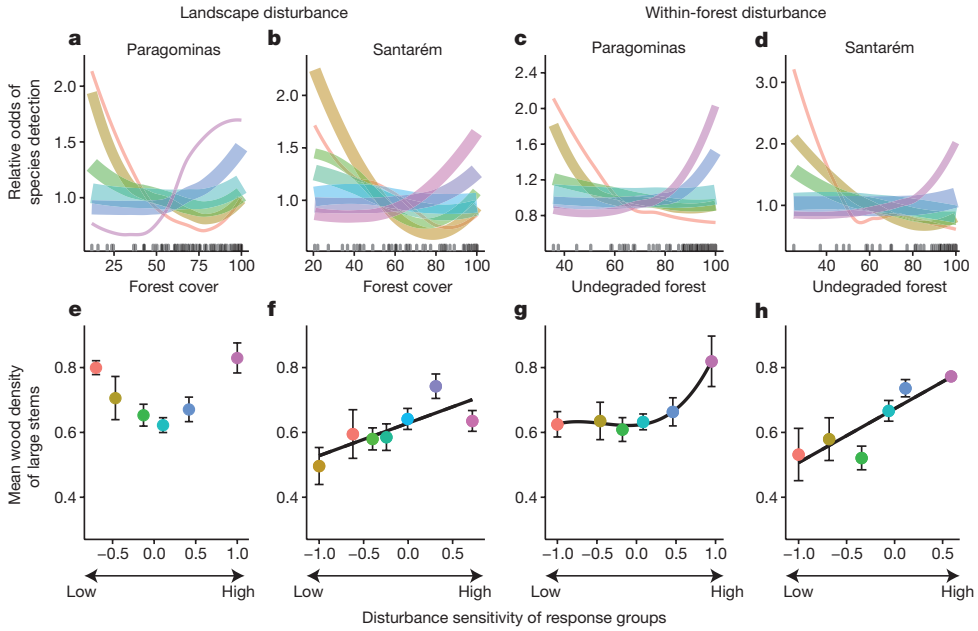


Figure 4 | Response of large-stemmed plants to disturbance. a–d, The odds of detecting species groups along gradients of landscape (**a**, **b**) and within-forest (**c**, **d**) disturbance in Paragominas (**a**, **c**) and Santarém (**b**, **d**) (see Methods). Species groups, shown by different coloured lines, are composed of species with similar disturbance responses (see Methods). Line thickness represents the relative size of the groups. **e–h**, Disturbance sensitivity of the species groups related to their mean wood density (g cm^{-3}). Error bars show s.e.m. Group colours correspond to groupings in **a–d**. Black lines show significant relationships ($P < 0.05$, F-test) (see Methods).

© 2016 Macmillan Publishers Limited. All rights reserved

remaining forests or work to actively reduce within-forest disturbance [6,21] (Extended Data Table 1).

Here we provide insights into the need for additional policies to reduce forest disturbance by examining the relative importance of landscape and within-forest disturbance on species distributions using Random Forests (see Methods). In ranking the importance of remotely sensed disturbance measures, we found that both forms of disturbance had significant additional effects on species' distributions, albeit with some region- and taxon-specific variation (Extended Data Figs 5–7 and Methods). We then used the measures of landscape and within-forest disturbance that were most frequently ranked highest to examine changes in taxon community structure, using Latent Trajectory Analysis to group species by their responses to disturbance (see Methods). Results showed a consistent and high level of community turnover from both forms of disturbance, with some species groups responding negatively and others positively (Figs 3 and 4 and Extended Data Fig. 8). These responses may explain the unimodal shape of the disturbance effect (Fig. 1b) because they are consistent with the loss of highly sensitive species at relatively low levels of forest disturbance and the dominance of more resistant taxa in the most disturbed forests. Finally, we linked species' response groups with life-history data available for birds and large-stemmed plants (see Methods). Both types of disturbance contributed to marked declines in species of high conservation and functional importance (birds with smaller range sizes[22,23] and plants with higher wood density[24–26], respectively) (Figs 3 and 4). These analyses almost certainly underestimate the adverse effects of disturbance because rare species, which are often most sensitive to human impacts in forest ecosystems[27], cannot be adequately modelled.

We provide compelling evidence that Amazonian conservation initiatives must address forest disturbance as well as deforestation. At its most stringent, Brazil's centrepiece environmental legislation, the Forest Code, mandates Amazonian landowners to maintain 80% of their primary forest cover. Our results show that even where this level of compliance is achieved, the primary forests of these landscapes may only retain 46–61% of their potential conservation value and are likely to have lost many species of high conservation and functional importance. These findings reinforce the need to reduce the effects of landscape fragmentation by zoning development activities, thereby ensuring the protection of large blocks of remaining forest in all biogeographic zones. Where deforestation has already occurred, further conservation losses can be minimised by preventing within-forest disturbance, aiding the recovery of already degraded forests, and investing in forest restoration to improve connectivity and buffer remnant forests from edge effects. Engendering change will require a mixture of incentive and regulatory-based measures to improve the sustainability of both forestry and farming practices. Crucially, because reducing forest disturbance requires coordinated efforts by many actors, interventions need to move beyond individual properties and address entire landscapes and regions. Such actions are urgently needed in the Amazon where logging operations are rapidly expanding across federal and state forests[28], wildfires are increasingly prevalent during more frequent and severe dry seasons[29], and the expansion of industrial agriculture, energy and mining threaten even strictly protected areas and indigenous lands[30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 March; accepted 17 May 2016.
Published online 29 June 2016.

1. Boucher, D., Elias, P., Faires, J. & Smith, S. Deforestation Success Stories: Tropical Nations Where Forest Protection and Reforestation Policies Have Worked. *Union of Concerned Scientists June 2014 Report* (2014).
2. Nepstad, D. *et al.* The end of deforestation in the Brazilian Amazon. *Science* **326,** 1350–1351 (2009).
3. Soares-Filho, B. S. *et al.* Modelling conservation in the Amazon basin. *Nature* **440,** 520–523 (2006).
4. Convention on Biological Diversity. *Strategic Plan for Biodiversity 2011–2020, Aichi Biodiversity Targets* https://www.cbd.int/sp/default.shtml (2015).
5. Legislative Database of the Food and Agricultural Organization of the United Nations (FAOLEX). Brazilian Environmental Law number 12.651 (25 March 2012).
6. Panfil, S. N. & Harvey, C. A. REDD+ and Biodiversity Conservation: A review of the biodiversity goals, monitoring methods and impacts of 80 REDD+ projects. *Conserv. Lett.* **9,** 143–150 (2015).
7. Aragão, L. E. O. C. & Shimabukuro, Y. E. The incidence of fire in Amazonian forests with implications for REDD. *Science* **328,** 1275–1278 (2010).
8. Burivalova, Z., Şekercioğlu, C. H. & Koh, L. P. Thresholds of logging intensity to maintain tropical forest biodiversity. *Curr. Biol.* **24,** 1893–1898 (2014).
9. Ewers, R. M. & Didham, R. K. Confounding factors in the detection of species responses to habitat fragmentation. *Biol. Rev. Camb. Philos. Soc.* **81,** 117–142 (2006).
10. Instituto Nacional de Pesquisas Espaciais (INPE). Projeto Prodes: Amazon deforestation database. Available at www.obt.inpe.br/prodes (2015).
11. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science* **342,** 850–853 (2013).
12. Sloan, S. & Sayer, J. Forest Resources Assessment of 2015 shows positive global trends but forest loss and degradation persist in poor tropical countries. *For. Ecol. Manage.* **352,** 134–145 (2015).
13. Barlow, J. & Peres, C. A. Avifaunal responses to single and recurrent wildfires in Amazonian forests. *Ecol. Appl.* **14,** 1358–1373 (2004).
14. Lewis, S. L., Edwards, D. P. & Galbraith, D. Increasing human dominance of tropical forests. *Science* **349,** 827–832 (2015).
15. Gibson, L. *et al.* Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* **478,** 378–381 (2011).
16. Malhi, Y., Gardner, T. A., Goldsmith, G. R., Silman, M. R. & Zelazowski, P. Tropical Forests in the Anthropocene. *Annu. Rev. Environ. Resour.* **39,** 125–159 (2014).
17. Morton, D. C., Le Page, Y., DeFries, R., Collatz, G. J. & Hurtt, G. C. Understorey fire frequency and the fate of burned forests in southern Amazonia. *Phil. Trans. R. Soc. B* **368,** 1–8 (2013).
18. Gardner, T. A. *et al.* A social and ecological assessment of tropical land uses at multiple scales: the Sustainable Amazon Network. *Phil. Tran. R. Soc. B* **368,** 20120166 (2013).
19. Berenguer, E. *et al.* A large-scale field assessment of carbon stocks in human-modified tropical forests. *Glob. Chang. Biol.* **20,** 3713–3726 (2014).
20. da Silva, J. M. C., Rylands, A. B. & Da Fonseca, G. A. B. The fate of the Amazonian areas of endemism. *Conserv. Biol.* **19,** 689–694 (2005).
21. International Union of Forest Research Organizations (IUFRO). *Understanding Relationships between Biodiversity, Carbon, Forests and People: The Key to Achieving REDD+ Objectives* (eds Parrotta, J. A., Wildburger, C. & Mansourian, S.) (2012).
22. Manne, L. L., Brooks, T. M. & Pimm, S. L. Relative risk of extinction of passerine birds on continents and islands. *Nature* **399,** 258–261 (1999).
23. Purvis, A., Gittleman, J. L., Cowlishaw, G. & Mace, G. M. Predicting extinction risk in declining species. *Proc. Biol. Sci.* **267,** 1947–1952 (2000).
24. Chave, J. *et al.* Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12,** 351–366 (2009).
25. Phillips, O. L. *et al.* Drought sensitivity of the Amazon rainforest. *Science* **323,** 1344–1347 (2009).
26. Baker, T. R. *et al.* Variation in wood density determines spatial patterns in Amazonian forest biomass. *Glob. Chang. Biol.* **10,** 545–562 (2004).
27. Banks-Leite, C. *et al.* Assessing the utility of statistical adjustments for imperfect detection in tropical conservation science. *J. Appl. Ecol.* **51,** 849–859 (2014).
28. Gestão de Florestas Públicas – Relatório 2015. Brasília: MMA/SFB Available at http://www.florestal.gov.br/publicacoes/instrumento-de-gestao (2015).
29. Chen, Y. *et al.* Forecasting fire season severity in South America using sea surface temperature anomalies. *Science* **334,** 787–791 (2011).
30. Ferreira, J. *et al.* Environment and Development. Brazil's environmental leadership at risk. *Science* **346,** 706–707 (2014).

Acknowledgements This work was supported by grants from Brazil (CNPq 574008/2008-0, 458022/2013-6, and 400640/2012-0; Embrapa SEG:02.08.06.005.00; The Nature Conservancy – Brasil; CAPES scholarships) the UK (Darwin Initiative 17-023; NE/F01614X/1; NE/G000816/1; NE/F015356/2; NE/I018123/1; NE/K016431/1), Formas 2013-1571, and Australian Research Council grant DP120100797. Institutional support was provided by the Herbário IAN in Belém, LBA in Santarém and FAPEMAT. R.M. and J.R.T. were supported by Australian Research Council grant DP120100797. This is paper no. 49 in the Sustainable Amazon Network series.

Author Contributions T.A.G., J.F. and J.B. designed the research with additional input from E.B., A.C.L., S.F.B.F., J.L., V.H.F.O., L.P., R.R.C.S., I.C.G.V., L.E.O.C.A. and R.P. E.B., A.C.L., V.H.F.O., R.R.C.S, R.F.B., J.F., R.C.O., N.G.M. R.C.S.V., J.L., J.M.S and F.Z.V. collected the field data or analysed biological or soil samples. G.D.L. analysed the data, with input from J.B., J.R.T., R.M., A.C.L. and T.A.G. S.F.B.F., R.A.B., T.M.C., C.M.S., S.S.N., J.V.S., A.V. and T.A.G. processed the remote sensing data. J.B., G.D.L., J.F., A.C.L., R.M., J.R.T. and T.A.G. wrote the manuscript, with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B. (josbarlow@gmail.com).

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Study regions. Pará is the second largest state in Brazil and a focal point for deforestation, accounting for 34% of all forest loss in the Brazilian Amazon between 1988 and 2015 (ref. 10). It holds exceptionally high biodiversity, with ~10% of the world's bird species and five of the eight major AoEs in Amazonia[20]. Within Pará, we focused on two geographically and biologically distinct regions: the municipalities of Paragominas and Santarém-Belterra-Mojuí dos Campos (abbreviated to Santarém) (Extended Data Fig. 1). These regions lie in different AoEs (Belém and Tapajós, respectively) and shared just 49% of our sampled taxa. Although they differ in their human colonization history[18], both retain >50% of their native forest cover.

Study design and biodiversity sampling. We divided each region into third- or fourth-order drainage catchments. In each region, 18 study catchments (32–61 km^2) were then distributed along forest cover gradients. We distributed study plots on *terra firme* in proportion to forest and non-forest cover at a density of approximately 1 plot per 4 km^2, resulting in 8–12 plots separated by ≥1.5 km in each catchment (Extended Data Fig. 1). Forest plots ($n = 234$) were distributed without previous knowledge of anthropogenic disturbance[18] and included primary forests (that is, under permanent forest cover; $n = 175$) and secondary forests recovering after agricultural abandonment ($n = 59$). Non-forest plots ($n = 133$) were predominantly located in pastures ($n = 76$) and mechanised agricultural lands ($n = 31$).

In total, 31 of the 36 catchments contained primary forest plots. In Paragominas and Santarém respectively, these included undisturbed (13 and 17), logged (44 and 26), burned (0 and 7) and logged and burned primary forests (44 and 24)[19]. Disturbance categories were based on field assessments of fire scars, charcoal and logging debris, and an analysis of canopy disturbance, deforestation and regrowth in time series satellite images (1988 to 2010)[18,19]. Plots in the undisturbed forest had no evidence of within-forest disturbance and, because they were located more than 2 km from edges in the largest forest blocks, had minimal landscape disturbance. Observations of hunting-sensitive large game birds, such as razor-billed curassow *Pauxi tuberosa* and trumpeters *Psophia* spp.[31,32], indicated low hunting pressure[33] in undisturbed plots.

Biodiversity surveys occurred during 2010 and 2011. The following descriptions apply to sampling at the plot level. Large and small stems: live trees and palms with ≥10 cm diameter at breast height were identified in 10 × 250 m plots. Smaller individuals (2–10 cm diameter) were sampled in five 5 × 20 m subplots (Extended Data Fig. 1). Liana diameters were measured at 1.3 m from the main root. Large- and small-stemmed plants were analysed separately because they may differ in their disturbance responses. Individuals were identified to species level by local parabotanists[19]. In total across all catchments, 175 plots and 825 subplots were sampled in primary forests. Birds: there were two repeat surveys of 15-min point counts at three sampling points (0, 150 and 300 m) (Extended Data Fig. 1). Sampling was undertaken between 15 min before dawn and 09:30. Lists of voucher sound-recordings and images are available for both regions[31,32]. In total across all catchments, 1,050 point counts were undertaken in primary forests. Dung beetles: sampled using pitfall traps (14 cm radius, 9 cm height) baited with 50 g of dung (80% pig and 20% human) and half filled with a killing solution (5% detergent and 2% salt). Traps were left for 48 h before inspection. Three traps were placed at the corners of a 3 m equilateral triangle, repeated at three sampling points (0, 150 and 300 m). In total across all catchments, 1,575 pitfall traps were set in primary forests (Extended Data Fig. 1).

Defining the biodiversity consequences of forest loss, landscape and within-forest disturbance. We limit the biodiversity consequences of forest loss to those that occur in deforested areas themselves, excluding any additional effects on remaining forests. Landscape disturbance then captures the combined edge, area and isolation effects that accompany the deforestation process. Within-forest disturbance refers to anthropogenic disturbance events that are not inevitable consequences of forest loss or land cover change, including wildfires, hunting and selective logging. Although often associated with landscape factors, such as distance from forest edge, within-forest disturbance can occur independently of changes in forest cover or landscape configuration.

Estimating the conservation value deficit. We used the sum of forest species presences in primary forest plots to measure the conservation value of a catchment. In practice, this means that if a forest species occurs on x plots within a catchment, the species contributes x to the catchment's conservation value. Total catchment conservation value is found by summing presences over all forest species. This measure is equivalent to mean species richness (per unit area) in primary forests multiplied by primary forest cover. In the absence of disturbance, conservation value should therefore respond linearly to forest cover, with slope equal to mean species density, d_e. We term the difference between this linear expectation and a catchment's observed conservation value as its conservation value deficit (CVD). We took a variety of approaches to calculating the CVD, reflecting different methods of defining forest species, weighting their importance, and calculating d_e.

Defining forest species. We restricted our analysis to 'forest species' to avoid attributing value to invasive and open-area species. We used three species classification filters: (i) an automatic filter defined forest species as those that occurred at least once in a primary forest plot, irrespective of the plot's disturbance history ($n = 1{,}621$ species); (ii) a high basal area (HBA) filter defined forest species to be those that occurred at least once in plots with a high average basal area (that is, greater than or equal to the lowest basal area recorded in undisturbed forests in each region) ($n = 1{,}290$); and (iii) a convex hull filter where we first applied a two-dimensional non-metric multidimensional scaling (MDS) to primary and secondary forest plots based on a stem-size classification (stress = 0.14), and then defined forest species to be those that occurred at least once in plots within the minimum convex hull of undisturbed primary forest plots on the MDS ($n = 1{,}140$).

Species conservation or functional importance. We used three approaches to weight species' importance. First, we assumed that all forest species had a value equal to 1. Second, we applied a linear weighting to birds according to their range size and plants according to their wood density. The bird species with the smallest range size was given a value of 1, and that with the largest range size was given a value of 0 (vice versa for plants and wood density), with all other species' values mapped linearly between these two points. Third, we squared the linear weighting to give an even higher relative value to species of highest conservation or functional importance.

There are many important life-history traits that correlate with species' conservation or functional importance. Our choices were based on a priori knowledge and the availability of data for diverse tropical taxa. For birds, we chose range size because it is the single most important predictor of threat status[23], especially among lowland passerines (which make up the majority of our sample) where it is inversely correlated with other important factors such as population density[22]. For plants, we chose wood density because it is the most important size-independent determinant of carbon storage within individual stems, a strong predictor of carbon stocks across the biome[24,26], and is also linked with other functional properties[24] including drought resistance[25]. Bird range sizes were extracted from the Birdlife Datazone (http://www.birdlife.org/datazone/index.html). Wood densities were adapted from the global wood density database[34], using the genus or family average where species or genus data were unavailable. Lianas were given a nominal value of 0.01.

As part of the broader sensitivity analysis we also undertook the same analysis described above for birds replacing species range size for species mean body size (body size data was also extracted from Birdlife Datazone). This analysis was undertaken to determine whether the population density of birds, which is strongly and inversely correlated with body size, significantly affected results. It did not: the median estimate of the disturbance impact decreased by just 0.5%, and we do not report the full results here.

Alternative undisturbed baselines. Estimating d_e (mean species density in undisturbed landscapes) requires species distribution data from catchments with no within-forest or landscape disturbance. As we do not have a set of such catchments in either region, we took three approaches to calculating d_e. The first two approaches rely on the least disturbed catchment in each region. In both Paragominas and Santarém, this reference catchment had minimal landscape disturbance (>99% primary forest). However, ground-based observations indicated that either selective logging or wildfire had affected at least 25% of the sampling plots within the reference catchments in both regions. We therefore calculated d_e as the mean species density over all plots in the reference catchment and, to correct for within-forest disturbance, as the mean density over only undisturbed reference catchment plots. Finally, to account for potential biases in underlying (natural) species distributions, we also calculated d_e using all undisturbed plots throughout each region ($n = 30$). This represents a more conservative estimate because it includes plots in catchments with less than 100% forest cover.

Selecting representative estimates of CVD. Combining the three forest species selection methods, the three species' weighting approaches, and the three estimates of d_e returns 27 estimates of the CVD. For all approaches, we determined the average CVD with respect to primary forest cover by modelling the catchments' summed presences with Poisson polynomial generalized linear models. We selected the best fitting model over all polynomials of degree up to cubics.

To express uncertainty over our estimates of the CVD, in the main text we present the median relationship between conservation value and forest cover along with the lower and upper bound range. We excluded from this range the estimate of d_e that included disturbed reference catchment plots because it is not reflective of species density in the absence of disturbance. For the purposes of comparison, we

© 2016 Macmillan Publishers Limited. All rights reserved

have included these results in Extended Data Fig. 4. The median, lower and upper bound estimates of the CVD were given by, respectively: the convex hull filter, linear species weighting, and undisturbed reference catchment plots; the convex hull filter, no species weighting, and all undisturbed plots; and the high basal area filter, exponential species weighting, and undisturbed reference catchment plots.

Adjusting for proportionality. Although the number of plots in catchments was proportional to forest cover, proportionality was not exact because the original distribution was based on the extent of both primary and secondary forests[18]. We therefore corrected sampling effort by calculating the factor required to make sampling proportional to primary forest cover in each catchment and scaled our estimates of conservation value accordingly. For each catchment i, this factor is given by p_i/t_i, where p_i is the proportion of catchment i that is primary forest and t_i is the number of primary forest transects in catchment i.

Extrapolating the CVD. To estimate disturbance impacts throughout Pará, we divided the state into grid cells approximately equal in size to our study catchments. We then used Brazil's 2010 Terraclass product[35] to determine the area of each cell that was deforested, first removing non-forested areas that were covered by water or tropical savannah. We then calculated each cell's conservation value by applying the median, lower and upper bound estimates of the CVD. The disturbance impact in forest loss equivalent terms for cell i is given by $p_i - (a_i - n_i)v_i$, where p_i, a_i, n_i and v_i are, respectively, the cell's primary forest extent, area, non-forest area and conservation value.

Linking landscape and within-forest disturbance with species distributions and traits. We investigated the importance of landscape and within-forest disturbance at the plot level rather than the catchment level because many disturbance drivers act at local scales[8,13]. Variables representing landscape and within-forest disturbance were based on the analysis of georeferenced 30 m resolution Landsat TM (Thematic Mapper) and eTM images from 1988 to 2010 in Paragominas and 1990–2010 in Santarém. These were complemented by covariates that represent natural variation in soil conditions, elevation and slope. A full description of the data are available elsewhere[18]. Variable abbreviations match those in Extended Data Figs 5–7.

Within-forest disturbance. We measured the cumulative extent of canopy disturbance[36] by calculating the percentage of the remaining primary forest in a 1 km buffer around each plot that had never been classified as disturbed (undisturbed primary forest, UPF). We also included two measures of the frequency of disturbance within plots: the number of times the plot was logged (NL) and the number of times the plot was burnt (NB) in visual inspections of satellite images or field observations.

Landscape disturbance. We used two landscape configuration measures: the density of forest-agriculture edges (ED) and the percentage of primary and secondary (>10 years old) forest cover (FC) in 1 km buffers around plots. We used two measures of landscape history[37]: the deforestation curvature profile (DC) and the land-use intensity profile (LI) in 500 m buffers around plots.

Natural environmental covariates. We used soil samples and digital elevation models to derive covariates reflecting natural conditions. Soil variables were based on average values from five 30 cm deep soil profiles in each plot, and include acidity (pH), clay content (Cl), and carbon stock (Ca). We applied a 100 m buffer around each plot in a digital elevation model to calculate mean plot elevation (El) and slope (Sl).

Linking landscape and within-forest disturbance with species distributions and traits. We used Random Forests (RF), a decision-tree classification methodology, to identify species that are well-modelled by our data and to rank the importance of individual variables in accounting for species distributions. RF was adapted for spatial autocorrelation within catchments using a modified 'residual autocorrelation' approach[38]. The fit of the RF models and their predictive performance was measured using area under receiver–operator curves (AUC)[39]. AUC evaluates the ability of models to correctly predict higher probability of occurrence where species are present than where they are absent. An AUC value of 1 indicates perfect discrimination; a value of 0.5 suggests predictions no better than random. We performed multiple cross-validations to evaluate model predictive performance. For each species, data from each study catchment were used in turn as test data for models built with data from the other catchments. The cross-validated AUC value, AUCcv, was calculated as the average AUC value over all cross-validation tests for each species. Species present on a minimum of three transects and with a summed AUCcv ≥ 0.6 over all variables were classified as well-modelled and included in the analyses (31% of species). The importance of a variable was measured as its mean AUCcv over all well-modelled species.

Models included the within-forest disturbance, landscape disturbance and natural environment covariates described above. Given multicollinearity, we selected two variables from each group using three variable-selection methods: (i) we selected variables that we hypothesized to have the greatest influence on species' presences (hypothesis-driven selection); (ii) we used principal component analysis (PCA) on the full set of variables in each group and selected the highest loaded variable on the first two principal axes (PCA selection); and (iii) we ran RF on the full set of variables and selected the two highest ranked in each group (step-wise selection). Results for each method are shown in Extended Data Figs 5–7.

Next, we used RF to determine species' partial responses along disturbance gradients (Figs 3 and 4 and Extended Data Fig. 8). These partial responses give the relative odds ($\exp(\text{logit}(p) - \text{mean}(\text{logit}(p))$, where p is the probability of species' presence and logit is $\ln(p/(1-p))$) of detecting each species along a single variable gradient, holding all other variables constant. For this analysis we selected the landscape and within-forest disturbance variables that were most frequently ranked highest in their group across the three variable selection methods.

We then used latent trajectory analysis (LTA), which groups species' partial responses into homogenous classes, to characterize the main types of response to the selected variables. We built models with up to eight classes and selected that with the lowest Bayesian information criterion (BIC) score. LTAs were carried out in R package 'lcmm' http://cran.r-project.org/web/packages/lcmm/lcmm.pdf. In Figs 3 and 4, we show the LOWESS smoothed response of each species class along the associated disturbance gradient, with bandwidth set to 0.75.

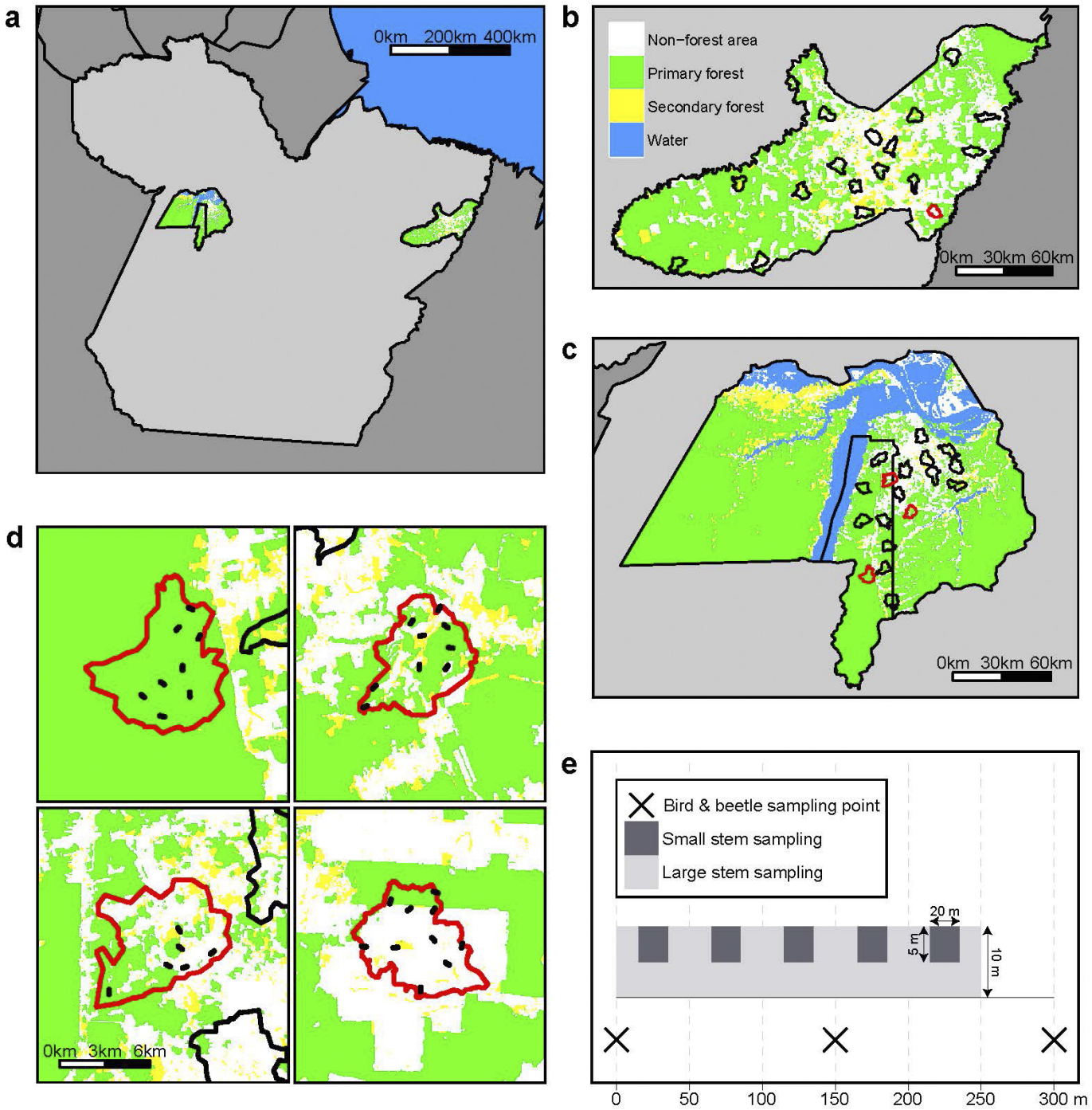
Finally, we investigated the relationship between the disturbance sensitivity of species classes, as determined by LTA, and species traits. To undertake this analysis, we defined a metric that represents the propensity of species classes to be detected along the variable gradients, which thus provides a measure of the sensitivity of the class to disturbance. The measure is:

$$h_c(x) = \int_m^u (x-m)d_c(x)dx - \int_l^m (m-x)d_c(x)dx$$

where m, l and u are, respectively, the gradient's mid-point and lower and upper bounds, and $d_c(x)$ is the relative odds of detecting species class c at point x on the gradient, as determined by RF. We scaled h_c to lie between $+1$ and -1. Values of h_c close to 1 indicate that species class c is much more likely to be detected at the maximum than minimum extreme of the gradient, values close to -1 indicate that species class c is much more likely to be found at the minimum than maximum extreme. Values near 0 indicate that species class c is equally likely to be detected at either extreme. We tested the relationship between h_c and species' traits by fitting polynomial models weighted by group size. In all cases, the response variable was the average value of the species trait over all species in each class. We investigated polynomial fits up cubics and selected that with the lowest BIC score.

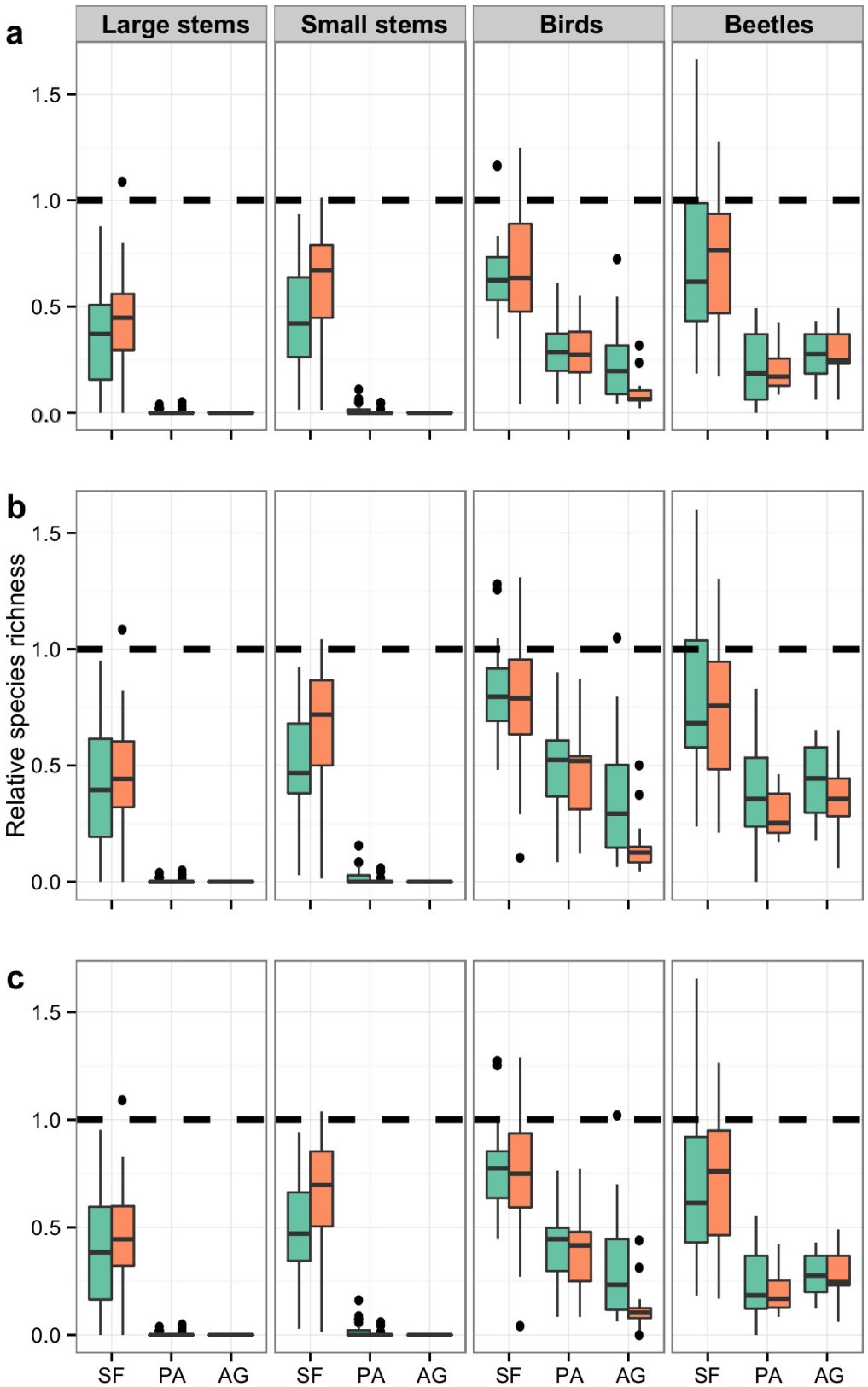
31. Lees, A. C. *et al.* One hundred and thirty-five years of avifaunal surveys around Santarem, central Brazilian Amazon. *Rev. Bras. Ornitol.* **21,** 16–57 (2013).
32. Lees, A. C. *et al.* Paragominas: a quantitative baseline inventory of an eastern Amazonian avifauna. *Rev. Bras. Ornitol.* **20,** 93–118 (2012).
33. Barrio, J. Hunting pressure on cracids (Cracidae: Aves) in forest concessions in Peru. *Rev. Peru. Biol.* **18,** 225–230 (2011).
34. Zanne A. E. *et al.* Data from: Towards a worldwide wood economics spectrum. Dryad Digital Repository. http://dx.doi.org/10.5061/dryad.234 (2009).
35. Instituto Nacional de Pesquisas Espaciais (INPE). Terraclass data 2010; available at http://www.Inpe.Br/cra/projetos_pesquisas/terraclass2010 (2013).
36. Souza, C. M. Jr. *et al.* Ten-year landsat classification of deforestation and forest degradation in the Brazilian Amazon. *Remote Sens.* **5,** 5493–5513 (2013).
37. Ferraz, S. F. D., Vettorazzi, C. A. & Theobald, D. M. Using indicators of deforestation and land-use dynamics to support conservation strategies: A case study of central Rondonia, Brazil. *For. Ecol. Manage.* **257,** 1586–1595 (2009).
38. Crase, B., Liedloff, A. C. & Wintle, B. A. A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography* **35,** 879–888 (2012).
39. Pearce, J. & Ferrier, S. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* **133,** 225–245 (2000).

© 2016 Macmillan Publishers Limited. All rights reserved



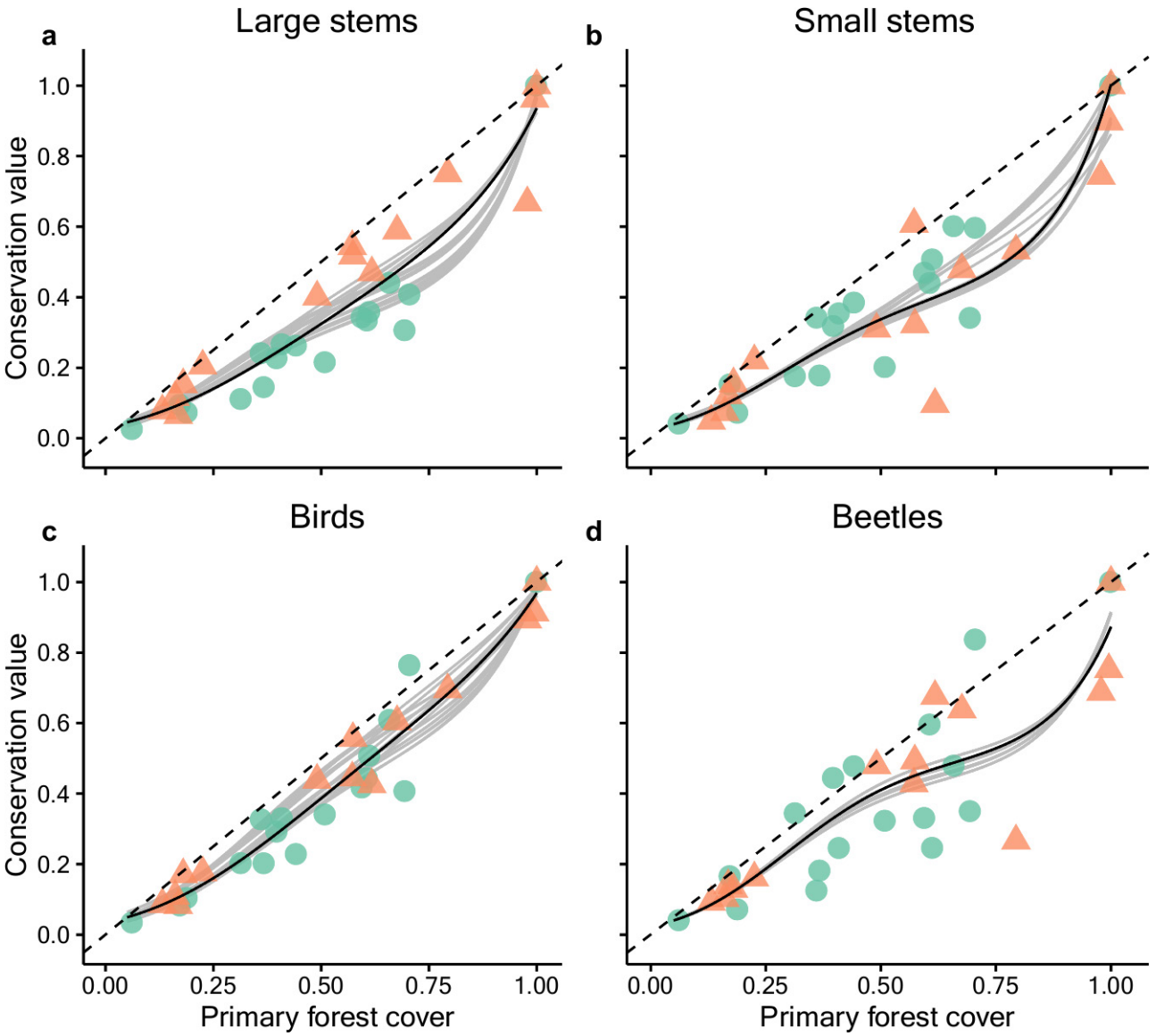
Extended Data Figure 1 | Study design. a, The location of Paragominas and Santarém within Pará. **b**, **c**, The distribution of study catchments ($n = 36$) within Paragominas and Santarém, respectively. **d**, The distribution of study plots ($n = 175$) in example catchments spanning the gradient of primary forest. Selected catchments are shown in red in **a** and **b**. **e**, Sampling design within each plot.

© 2016 Macmillan Publishers Limited. All rights reserved



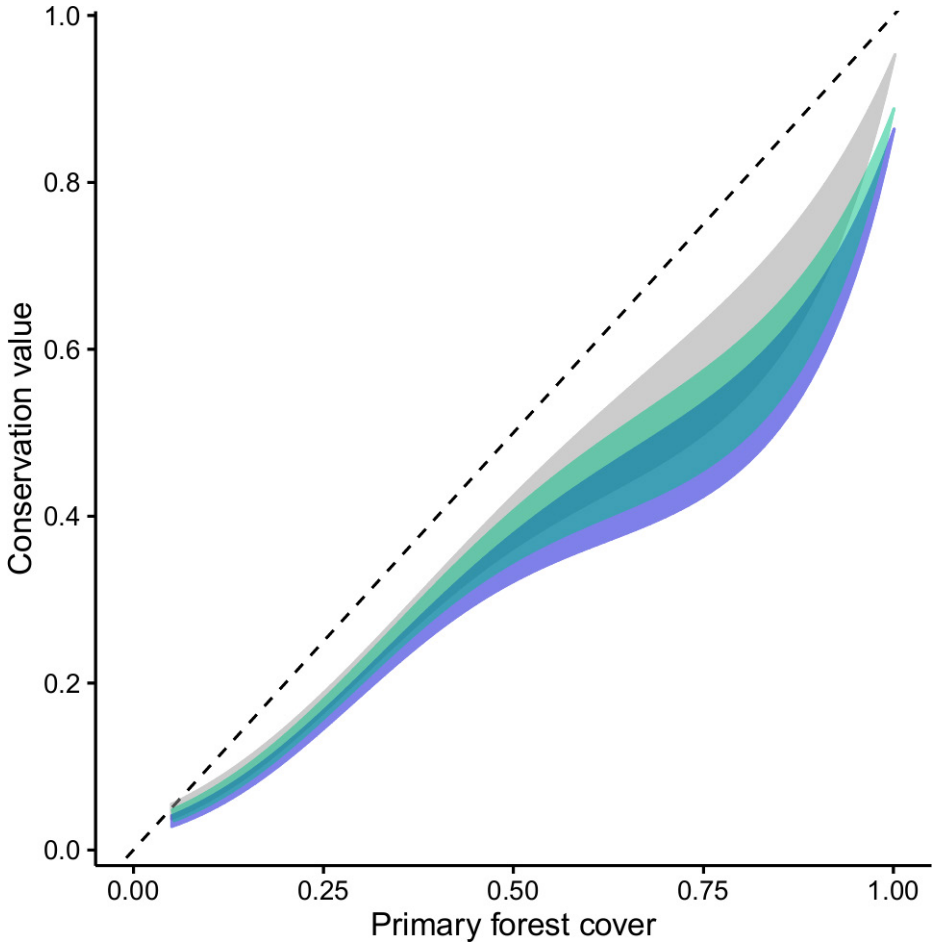
Extended Data Figure 2 | Richness of forest species. a–c, The richness of forest species in secondary forests (SF), pastures (PA), and mechanised agricultural lands (AG) relative to the average richness of forest species in all undisturbed and disturbed primary forests (dashed line) in Paragominas (green) and Santarém (orange). Panels show the convex hull (**a**), automatic (**b**) and high basal area filters (**c**) used to classify forest species (see Methods).

© 2016 Macmillan Publishers Limited. All rights reserved



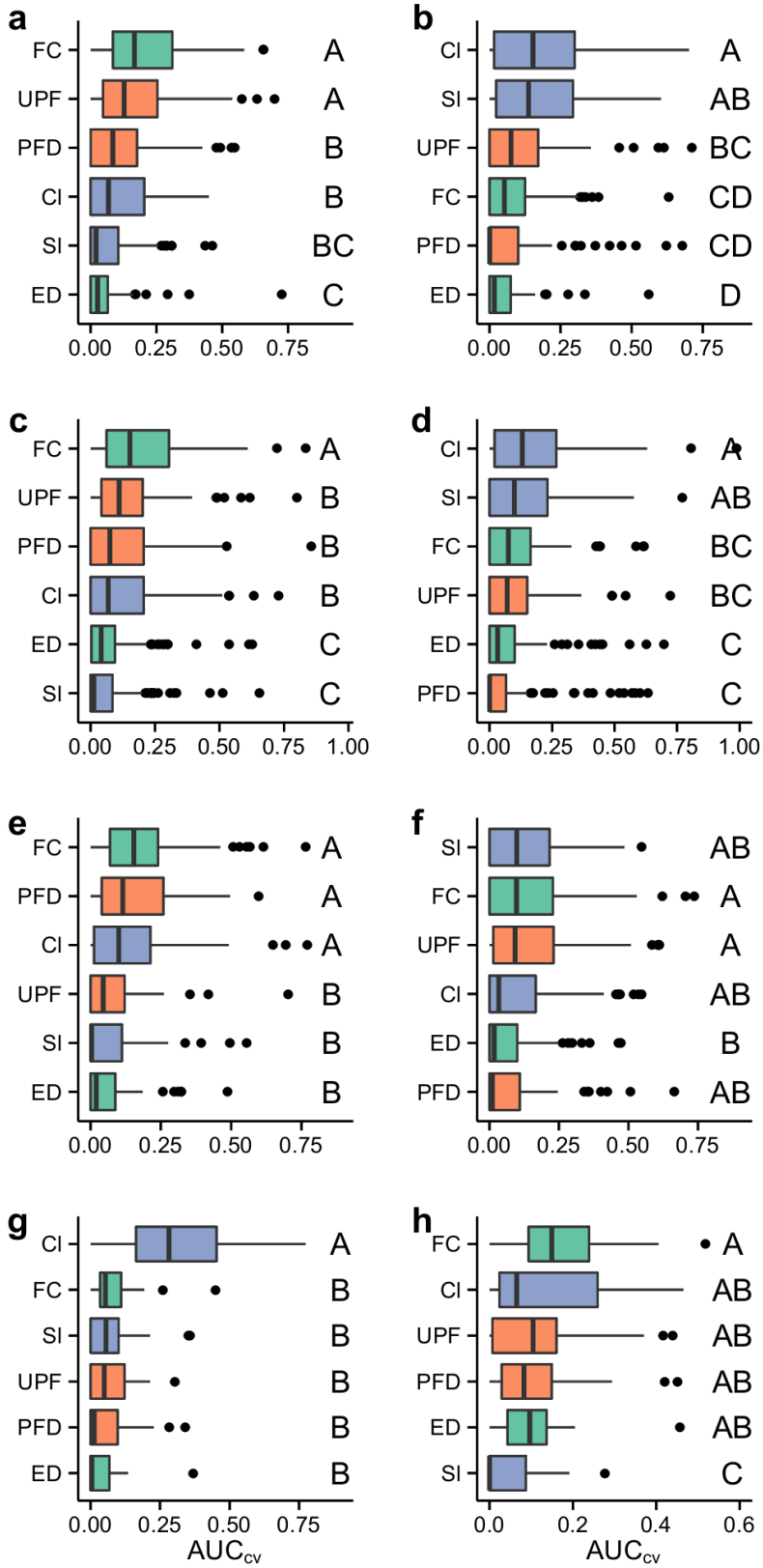
Extended Data Figure 3 | Conservation value of primary forests measured by individual taxa. a–d, Estimates of conservation value in the Paragominas (circles) and Santarém (triangles) study regions from large-stemmed plants (**a**) small-stemmed plants (**b**) birds (**c**) and dung beetles (**d**). Dashed lines show expectations without disturbance. Grey lines show all regressions, with the black solid line showing the median response (see Methods). Values were standardized across study regions and taxa. There was no significant difference between taxa in the median estimate ($F_{3;117} = 1.36$, $P = 0.26$, ANCOVA).

© 2016 Macmillan Publishers Limited. All rights reserved



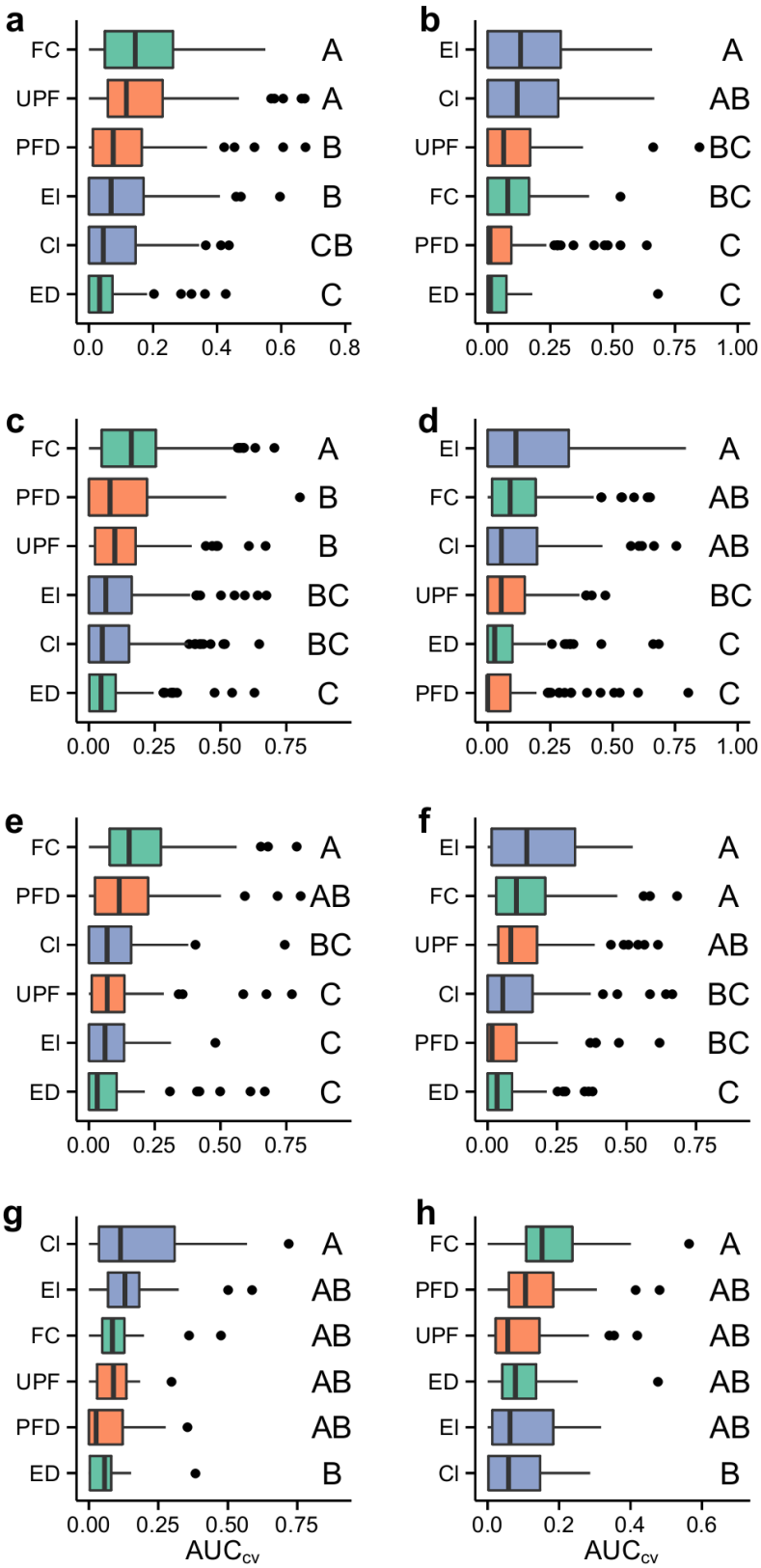
Extended Data Figure 4 | Range of conservation value estimates using three alternative sets of reference plots. Mean species density (d_e) is measured by: all disturbed and undisturbed plots in the least disturbed reference catchments (grey shaded region), all undisturbed plots throughout a region (green shaded region), and undisturbed plots in the reference catchments (purple shaded region). See Methods for details.

© 2016 Macmillan Publishers Limited. All rights reserved



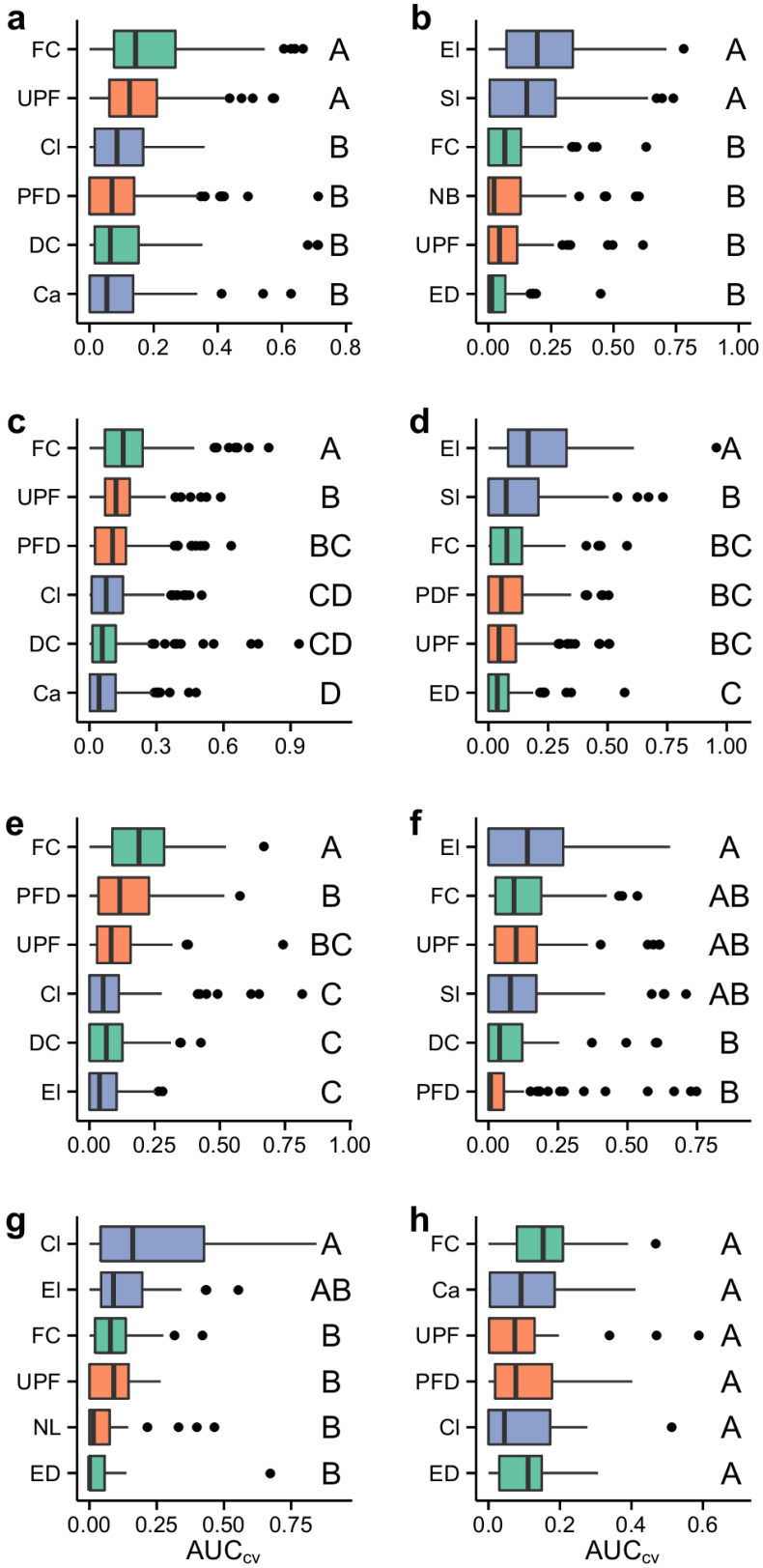
Extended Data Figure 5 | The importance of hypothesis selected variables. **a**–**h**, Species AUCcv values for each variable in Paragominas (**a**, **c**, **e**, **g**) and Santarém (**b**, **d**, **f**, **h**) for large-stemmed plants (**a**, **b**), small-stemmed plants (**c**, **d**), birds (**e**, **f**) and beetles (**g**, **h**). Variable importance was measured by the mean AUCcv over all well-modelled species (see Methods). Variable colours denote group membership: green, orange and blue represent landscape disturbance, within-forest disturbance and natural variables, respectively (see Methods for variable descriptions). Letters show the results for multiple pair-wise comparisons of group means using Tukey's range test. Variables which do not share a letter have significantly different mean importance ($P < 0.05$).

© 2016 Macmillan Publishers Limited. All rights reserved



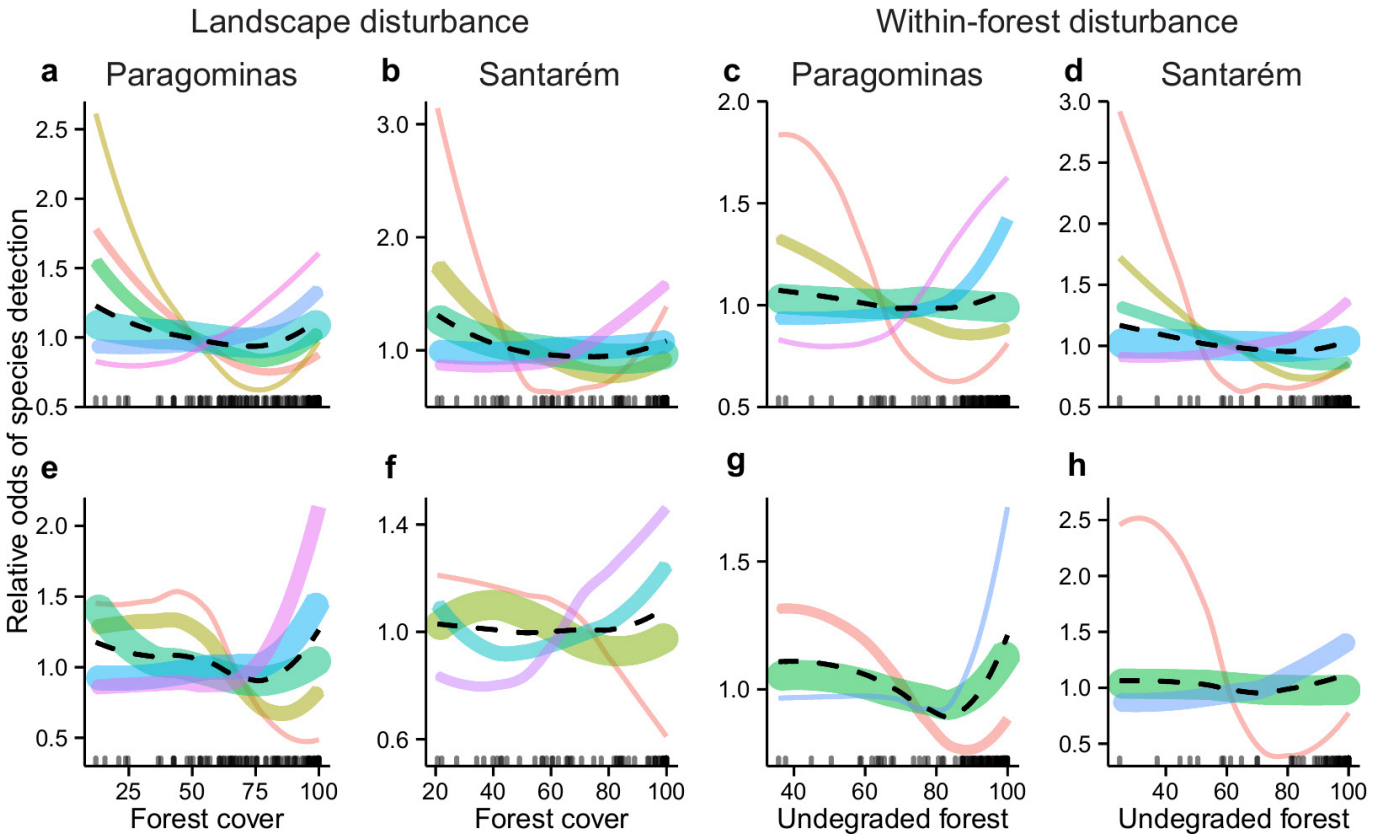
Extended Data Figure 6 | The importance of PCA selected variables. **a**–**h**, Species' AUCcv values for each variable in Paragominas (**a**, **c**, **e**, **g**) and Santarém (**b**, **d**, **f**, **h**) for large-stemmed plants (**a**, **b**), small-stemmed plants (**c**, **d**), birds (**e**, **f**) and beetles (**g**, **h**). Variable importance was measured by the mean AUCcv over all well-modelled species (see Methods). Variable colours denote group membership: green, orange and blue represent landscape disturbance, within-forest disturbance and natural variables, respectively (see Methods for variable descriptions). Letters show the results for multiple pair-wise comparisons of group means using Tukey's range test. Variables which do not share a letter have significantly different mean importance ($P < 0.05$).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | The importance of step-wise selected variables. **a**–**h**, Species' AUCcv values for each variable in Paragominas (**a**, **c**, **e**, **g**) and Santarém (**b**, **d**, **f**, **h**) for large-stemmed plants (**a**, **b**), small-stemmed plants (**c**, **d**), birds (**e**, **f**) and beetles (**g**, **h**). Variable importance is measured by the mean AUCcv over all well-modelled species (see Methods). Variable colours denote group membership: green, orange and blue represent landscape disturbance, within-forest disturbance and natural variables, respectively (see Methods for variable descriptions). Letters show the results for multiple pair-wise comparisons of group means using Tukey's range test. Variables which do not share a letter have significantly different mean importance ($P < 0.05$).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | Responses of small-stemmed plants and dung beetles to disturbance. a–h, The odds of detecting small-stemmed plants (**a**–**d**) and dung beetles (**e**–**h**) species groups along gradients of landscape disturbance (**a**, **b**, **e**, **f**) and within-forest disturbance (**c**, **d**, **g**, **h**) in Paragominas (**a**, **c**, **e**, **g**) and Santarém (**b**, **d**, **f**, **h**) (see Methods). Species groups, shown by different coloured lines, are composed of species with similar disturbance responses (see Methods). Line thickness represents the relative size of the groups.

© 2016 Macmillan Publishers Limited. All rights reserved

Extetnded Data Table 1 | Policy interventions used to reduce deforestation and their effect on disturbance

Policy intervention	*Direct effects on reducing landscape disturbance*	*Direct effects on reducing within-forest disturbance*
Protected areas (IUCN classes I-IV)	Positive if there is no leakage of deforestation	Positive if park management is effective and leakage of logging is avoided
Sustainable-use reserves (IUCN class VI)	Positive if there is no leakage of deforestation	Positive where more sustainable approaches replace conventional approaches, and if leakage of logging is avoided Negative if forest-use is incentivised in areas that would not otherwise be disturbed
Legal stipulation to maintain forest cover on private lands	Positive, but there is no stipulation to consider landscape configuration	No likely impact without additional measures
Agricultural intensification on deforested lands	Positive if this prevents further forest loss Negative if increased profits encourage further land-use change Negative if the matrix becomes more hostile to forest species, increasing isolation	Positive if reduced fire use in agriculture prevents wildfires No likely impact on selective logging or hunting Negative if there are new spillover effects from agriculture, such as deposition of nutrients and pesticides
Industrial and community based reduced impact logging	Positive if economic returns protect forests from clearance Negative when new roads and logging patios increase edge-effects and isolation	Positive if more sustainable approaches replace conventional logging Negative when logging is incentivised in undisturbed forests
Protecting forests through moratoria & certification	Positive if this prevents further forest loss and there is no leakage of deforestation	No likely impact without additional measures

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 2 | Forest loss and disturbance in Pará and its areas of endemism

Region	Area	Forest area	Forest loss	Disturbance	Relative (%)
All land (a)					
Pará state	1,259,916	1,141,659	245,288	171,695 (130,827-187,328)	70 (53-76)
Belém AoE	138,351	131,637	80,861	15,431 (10,904-18,743)	19 (13-23)
Guiana AoE	273,692	246,802	12,290	37,044 (29,134-37,124)	301 (237-302)
Rondônia AoE	66,222	56,876	5,728	9,366 (7,227-10,103)	163 (126-177)
Tapajós AoE	418,201	388,738	37,783	65,221 (50,117-70,951)	173 (133-188)
Xingu AoE	321,304	294,543	109,687	40,502 (30,173-45,918)	37 (28-42)
Private lands + sustainable use reserves (b)					
Pará state	918,694	820,636	242,578	122,881 (92,144-139,033)	51 (38-57)
Belém AoE	136,405	129,691	80,432	15,099 (10,652-18,358)	19 (13-23)
Guiana AoE	157,288	137,468	12,084	20,879 (16,278-21,497)	173 (135-178)
Rondônia AoE	55,109	45,832	5,559	7,631 (5,859-8,379)	137 (105-151)
Tapajós AoE	271,761	250,309	36,466	43,980 (33,303-49,795)	121 (91-137)
Xingu AoE	253,884	232,681	109,029	30,891 (22,572-36,260)	28 (21-33)
Private lands (c)					
Pará state	530,931	490,200	230,293	68,694 (49,731-82,024)	30 (22-36)
Belém AoE	129,570	125,209	79,072	14,324 (10,077-17,435)	18 (13-22)
Guiana AoE	55,068	42,069	11,615	6,567 (4,912-7,491)	57 (42-64)
Rondônia AoE	33,676	28,033	4,551	4,692 (3,574-5,248)	103 (79-115)
Tapajós AoE	125,089	118,409	31,401	21,982 (16,126-26,436)	70 (51-84)
Xingu AoE	196,824	182,752	106,276	22,159 (15,783-26,675)	21 (15-25)

a–**c**, The loss of primary forest conservation value from forest loss and forest disturbance in forest loss-equivalent terms across ~50 km^2 cells covering all land in Pará (**a**), private lands and sustainable use reserves only (**b**), and private lands only (**c**). Disturbance losses are calculated using the median estimate of conservation value with the lower and upper bound range in parentheses (see Methods). Area is the total area of the region in km^2. Forest area gives the area of the region that was or is primary forest cover in km^2. Forest loss gives the total loss of primary forest in km^2. Disturbance gives the loss of conservation value due to disturbance in km^2. Relative gives the disturbance-mediated loss of conservation value relative to that from forest loss.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18621

Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases

Ying-Nan P. Chen[1], Matthew J. LaMarche[1], Ho Man Chan[1], Peter Fekkes[1], Jorge Garcia-Fortanet[1], Michael G. Acker[1], Brandon Antonakos[1], Christine Hiu-Tung Chen[1], Zhouliang Chen[1], Vesselina G. Cooke[1], Jason R. Dobson[1], Zhan Deng[1], Feng Fei[1], Brant Firestone[1], Michelle Fodor[1], Cary Fridrich[1], Hui Gao[1], Denise Grunenfelder[1], Huai-Xiang Hao[1], Jaison Jacob[1], Samuel Ho[1], Kathy Hsiao[1], Zhao B. Kang[1], Rajesh Karki[1], Mitsunori Kato[1], Jay Larrow[1], Laura R. La Bonte[1], Francois Lenoir[1], Gang Liu[1], Shumei Liu[1], Dyuti Majumdar[1], Matthew J. Meyer[1], Mark Palermo[1], Lawrence Perez[1], Minying Pu[1], Edmund Price[1], Christopher Quinn[1], Subarna Shakya[1], Michael D. Shultz[1], Joanna Slisz[1], Kavitha Venkatesan[1], Ping Wang[1], Markus Warmuth[1], Sarah Williams[1], Guizhi Yang[1], Jing Yuan[1], Ji-Hu Zhang[1], Ping Zhu[1], Timothy Ramsey[1], Nicholas J. Keen[1], William R. Sellers[1], Travis Stams[1] & Pascal D. Fortin[1]

The non-receptor protein tyrosine phosphatase SHP2, encoded by *PTPN11*, has an important role in signal transduction downstream of growth factor receptor signalling and was the first reported oncogenic tyrosine phosphatase[1]. Activating mutations of SHP2 have been associated with developmental pathologies such as Noonan syndrome and are found in multiple cancer types, including leukaemia, lung and breast cancer and neuroblastoma[1–5]. SHP2 is ubiquitously expressed and regulates cell survival and proliferation primarily through activation of the RAS–ERK signalling pathway[2,3]. It is also a key mediator of the programmed cell death 1 (PD-1) and B- and T-lymphocyte attenuator (BTLA) immune checkpoint pathways[6,7]. Reduction of SHP2 activity suppresses tumour cell growth and is a potential target of cancer therapy[8,9]. Here we report the discovery of a highly potent ($IC_{50} = 0.071\,\mu M$), selective and orally bioavailable small-molecule SHP2 inhibitor, SHP099, that stabilizes SHP2 in an auto-inhibited conformation. SHP099 concurrently binds to the interface of the N-terminal SH2, C-terminal SH2, and protein tyrosine phosphatase domains, thus inhibiting SHP2 activity through an allosteric mechanism. SHP099 suppresses RAS–ERK signalling to inhibit the proliferation of receptor-tyrosine-kinase-driven human cancer cells *in vitro* and is efficacious in mouse tumour xenograft models. Together, these data demonstrate that pharmacological inhibition of SHP2 is a valid therapeutic approach for the treatment of cancers.

To discover new cancer therapeutic targets, a deep-coverage, pooled short hairpin RNA (shRNA) library targeting 7,500 genes with 20 shRNAs per gene was screened across a panel of 250 cell lines from the Cancer Cell Line Encyclopedia (CCLE) (ref. 10 and Schlabach *et al.*, manuscript in preparation). An unbiased correlation analysis was performed and revealed that cell lines sensitive to SHP2 depletion are most sensitive to EGFR depletion ($P = 4.10 \times 10^{-8}$). When a subset of cell lines dependent on known receptor tyrosine kinases (RTKs) (such as EGFR, ERBB2, c-MET, and FLT3) and FRS2-dependent lines were considered as a class, a marked correlation emerged with sensitivity to SHP2 depletion (Fig. 1a and Extended Data Fig. 1, Fisher's exact $P < 4.45 \times 10^{-14}$). These findings provide a robust cross-validation of reports that RTK-driven cancer cells depend on SHP2 for survival[8,9]. Conversely, cell lines that were sensitive to KRAS, NRAS or BRAF depletion were refractory to SHP2 downregulation (Fig. 1b and Extended Data Fig. 1, Fisher's exact $P < 7.90 \times 10^{-5}$). To validate these findings further, doxycycline (dox)-inducible SHP2 shRNAs were introduced into cancer cells lines with distinct RTK alterations, including amplification of *EGFR* (MDA-MB-468, KYSE520), *ERBB2* (NCI-H2170), *FGFR2* (SUM52 and KATO III), and *EML4-ALK* translocation (NCI-H2228). Consistent with the shRNA screening data, SHP2 depletion led to marked inhibition of colony formation in each of these RTK-dependent cancer cells (Extended Data Fig. 1a). Importantly, this was specific to RTK-dependent cell lines, as *BRAF*- and *KRAS*-mutated cells (A2058 and MDA-MB-231) showed no growth effect upon SHP2 depletion (Extended Data Fig. 1b). To evaluate the importance of SHP2 catalytic activity for the growth of sensitive cell lines, a complementation experiment was conducted by re-expressing shRNA-resistant alleles of wild-type SHP2 or the catalytically inactive $SHP2^{C459S}$ variant in MDA-MB-468 cells. SHP2 depletion inhibited the growth of MDA-MB-468 accompanied by reduced p-ERK levels (Fig. 1c, Extended Data Fig. 1a, c). Upon dox treatment, wild-type SHP2 and $SHP2^{C459S}$ were expressed at similar levels. Expression of wild-type SHP2, but not $SHP2^{C459S}$ restored p-ERK levels and cell growth (Fig. 1c, Extended Data Fig. 1c). Similar results were obtained with SUM52 cells (Extended Data Fig. 2). Therefore, SHP2 phosphatase activity is required for p-ERK activation and maintenance of cell growth in RTK-driven cancers.

On the basis of the shRNA screening results, we hypothesized that cells with constitutively activated RAS signalling would be insensitive to SHP2 inhibition. To test this hypothesis, SHP2-dependent SUM52 cells were transduced with a lentivirus carrying the $KRAS^{G12V}$ oncogene. Expression of $KRAS^{G12V}$ restored p-ERK levels and rendered these cells insensitive to SHP2 knockdown (Fig. 1d, Extended Data Fig. 1d). Furthermore, SHP2 depletion had no impact on cell growth and proliferation in MDA-MB-231 ($KRAS^{G13D}$) or A2058 ($BRAF^{V600E}$) cells (Extended Data Fig. 1b). These data strongly suggest that cancer cells carrying oncogenic *RAS/RAF* mutations will be refractory to SHP2 inhibition.

Efforts to discover small molecule therapeutics targeting protein tyrosine phosphatases (PTPs) have been challenged by the highly solvated and polar nature of the catalytic site, as exemplified by the SHP2 PTP domain[11–15]. To discover novel modes of phosphatase inhibition, we developed screening strategies aimed at identifying SHP2 allosteric inhibitors. SHP2 is activated by peptides and proteins containing appropriately spaced phospho-tyrosine residues that bind the N-terminal and C-terminal SH2 domains (denoted as N-SH2 and C-SH2, respectively) in a bidentate manner, releasing the auto-inhibitory interface and making the active site available for substrate recognition and turnover[16,17]. To discover inhibitors that could take advantage of this natural regulatory mechanism and lock SHP2 in an auto-inhibited conformation (Fig. 2a), a diverse library of 100,000 compounds were screened at 20 μM against SHP2 (residues 1–525) that was partially activated using 0.5 μM of a bisphosphorylated IRS-1

[1]Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

© 2016 Macmillan Publishers Limited. All rights reserved

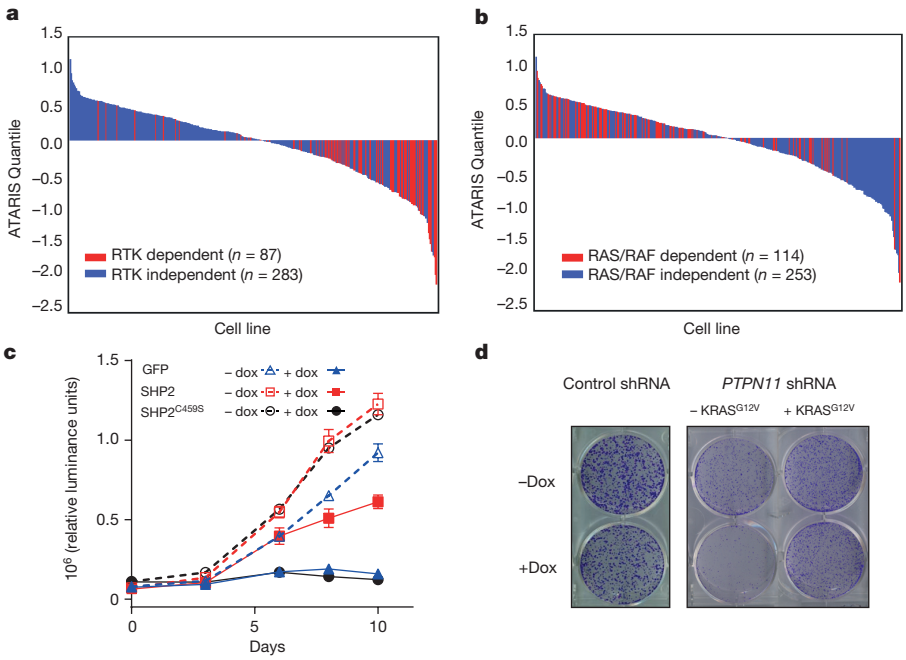


Figure 1 | Genetic validation of SHP2. **a**, Waterfall plot showing the ATARIS Quantile score for SHP2 shRNAs coloured by the effect of RTK shRNA knockdown (ATARIS score <-1.0) in 370 cell lines. **b**, Waterfall plot showing the ATARIS Quantile score for SHP2 shRNAs coloured by the effect of shRNA knockdown of KRAS, NRAS or BRAF (ATARIS score <-1.0) in 367 cell lines. **c**, Cell proliferation in MDA-MB-468 SHP2 knockdown cells stably expressing GFP, wild-type haemagglutinin (HA)-tagged SHP2 or HA–SHP2^{C459S}. Cells were treated with dox (100 ng ml^{-1}) and cell growth was measured by CellTiter-Glo assay at the indicated times. Data presented as mean ± s.d. ($n = 3$). **d**, Colony formation of SHP2-depleted SUM52 cells stably expressing vector control or HA–KRAS^{G12V}. In **c** and **d**, dox treatment induces depletion of endogenous SHP2 protein and simultaneous expression of the exogenous proteins GFP, wild-type HA–SHP2, HA–SHP2^{C459S} or HA–KRAS^{G12V}.

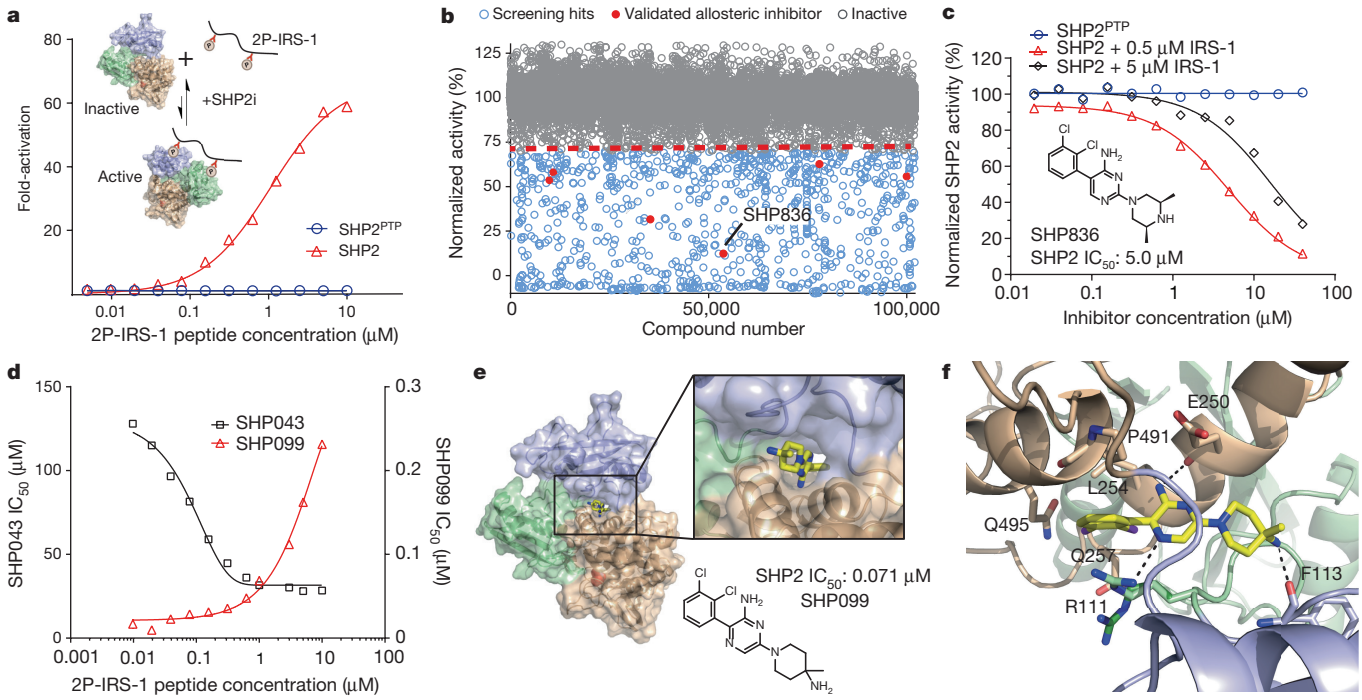


Figure 2 | Discovery of a SHP2 allosteric inhibitor. **a**, Schematic of SHP2 allosteric activation by 2P-IRS-1, highlighting an allosteric inhibitor that blocks the activation of SHP2 via the enrichment of its auto-inhibited conformer and dose-dependent activation of SHP2 by 2P-IRS-1 peptide. SHP2, SHP2^{PTP} and a dimethyl sulfoxide (DMSO) control were incubated with increasing concentrations of 2P-IRS-1 peptide. Biochemical activity was monitored using the DiFMUP (6,8-difluoro-4-methylumbelliferyl phosphate) substrate and normalized against the basal activity determined for each condition in the absence of 2P-IRS-1. SHP2i, SHP2 inhibitor. **b**, Primary screen was performed using a 100,000-molecule library. SHP2 was screened in the presence of 0.5 μM 2P-IRS-1 and 20 μM of each compound. The red dotted line represents the 30% inhibition threshold. The red circles represent the six validated allosteric inhibitors. SHP836 is labelled, and inhibited SHP2 activity by 87.6% The Z' factor determined for the screen was >0.9. **c**, Biochemical assay fingerprint observed with allosteric inhibitor SHP836. Inhibition of SHP2^{PTP} and SHP2 in the presence of 0.5 μM 2P-IRS-1 and 5 μM 2P-IRS-1. **d**, SHP2 inhibition by SHP099 and PTP active-site inhibitor SHP043 in the presence of various 2P-IRS-1 concentrations. **e**, Chemical structure of SHP099 and X-ray crystal structure of SHP2 in complex with SHP099 (PDB accession number 5EHR). Surface representation of SHP2 in complex with SHP099 bound in the central tunnel formed at the interface of the three domains (green, N-SH2; blue, C-SH2; tan, PTP domain). SHP2 is in the inactive conformation with the N-terminal SH2 domain fully occluding the entry of substrate to the active-site cysteine (shown in red). **f**, Key interactions between SHP099 and all three domains of SHP2 highlighted, including hydrogen bonds with Arg111 (N-SH2), Phe113 (C-SH2), and Glu250 (PTP). Data points along the line in **a**, **c** and **d** represent the mean of two replicate values.

© 2016 Macmillan Publishers Limited. All rights reserved

peptide (2P-IRS-1). Nine hundred compounds were found to inhibit the enzyme by 30% or more (Fig. 2b) and further profiled in three distinct biochemical assays: (1) using a truncated form of SHP2 with the PTP domain only ($SHP2^{PTP}$), or SHP2 assayed with (2) partially and (3) fully activating levels of 2P-IRS-1. Compounds that inhibited only the phosphatase domain were deprioritized to enrich for potential allosteric inhibitors. Six compounds, exemplified by SHP836, demonstrated no inhibition against $SHP2^{PTP}$, moderate inhibitory activity ($IC_{50} = 5–50\,\mu M$) against SHP2 activated by 0.5 μM 2P-IRS-1, and reduced inhibitory activity in the presence of a higher concentration of 2P-IRS-1 (Fig. 2c). SHP836 was further optimized to SHP099, yielding a >70-fold improvement in biochemical potency to $IC_{50} = 0.071\,\mu M$ (manuscript in preparation). Furthermore, SHP099 showed no detectable activity against a panel of 21 phosphatases and 66 kinases[18] (Extended Data Tables 1 and 2), and only had modest activity against 5HT3 when profiled against a preclinical safety pharmacology panel representing 49 common adverse drug reaction targets[19] (Extended Data Table 3). Importantly, SHP099 showed no activity against SHP1, the closest homologue of SHP2 sharing 61% amino acid sequence identity, supporting its high degree of target selectivity.

To understand the mechanism of inhibition, we determined the effect of the 2P-IRS-1 peptide on the potency of SHP099 and compared it to the active site inhibitor SHP043, stemming from a previously reported class of PTP1B inhibitors[20] (Fig. 2d). An increase in 2P-IRS-1 peptide correlated with a tenfold decrease in SHP099 potency, as opposed to a sixfold increase in SHP043 potency across the same range of 2P-IRS-1 peptide concentrations (Fig. 2d). These data suggest that SHP099 is interfering with the 2P-IRS-1-driven activation of SHP2, and that the active site of the PTP is more readily accessible for SHP043 binding in the activated conformation of SHP2. In addition, isothermal titration calorimetry revealed that SHP099 bound to SHP2 with 1:1 stoichiometry with a measured dissociation constant of 0.073 μM ($K = 1.38 \times 10^7\,M^{-1}$, Extended Data Fig. 3a).

To distinguish further the mechanism of inhibition from catalytic site inhibitors, we solved the crystal structure of SHP099 in complex with SHP2 (resolution, 1.7Å) (Extended Data Table 4). Here, SHP2 was found in the same auto-inhibited, inactive conformation as the reported apo-SHP2 structure[17], with the N-terminal SH2 domain blocking the active site. Our structure revealed SHP099 bound to the central tunnel formed at the interface of the N-SH2, C-SH2, and PTP domains (Fig. 2e) and interactions between SHP099 and all three domains of SHP2, strongly suggesting that SHP099 inhibits the catalytic activity through stabilization of the inactive conformation of the enzyme. Key interactions include hydrogen bonds with Arg111 and Phe113 located on the linker between the N-SH2 and C-SH2 domains, as well as Glu250 from the PTP domain. Additionally, the dichlorophenyl group of SHP099 makes extensive hydrophobic interactions with the sidechains of Leu254, Gln257, Pro491, and Gln495 of the PTP domain (Fig. 2f). In the homologue SHP1, repositioning of the linker between the two SH2 domains would remove key SHP099 interactions (highlighted by residue Arg109 in SHP1 and Arg111 in SHP2; Extended Data Fig. 3b–d), and would yield a significantly larger central tunnel with an estimated volume of 1,012 $Å^3$ compared to 464 $Å^3$ for SHP2 and to the 262 $Å^3$ volume of SHP099. These observations probably explain the selectivity of SHP099 for SHP2 over SHP1.

To determine whether SHP099 was capable of cellular SHP2 inhibition, cells were treated with increasing concentrations of SHP099. SHP099 inhibited p-ERK with an IC_{50} of ~0.25 μM in SHP2-dependent MDA-MB-468 and KYSE520 cells, but not in A2058 cells (Fig. 3a). No effect was observed on p-AKT levels across the same cells (Extended Data Fig. 4a). The inhibition of p-ERK was consistent with the growth inhibition observed in a colony-formation assay (Extended Data Fig. 4b). The inhibition of KYSE520, MDA-MB-468 and A2058 cells by SHP099 was also assessed in a cell proliferation assay and extended to three additional SHP2-dependent haematopoietic cell lines, MV-411, MOLM-13 and Kasumi-1, resulting in the

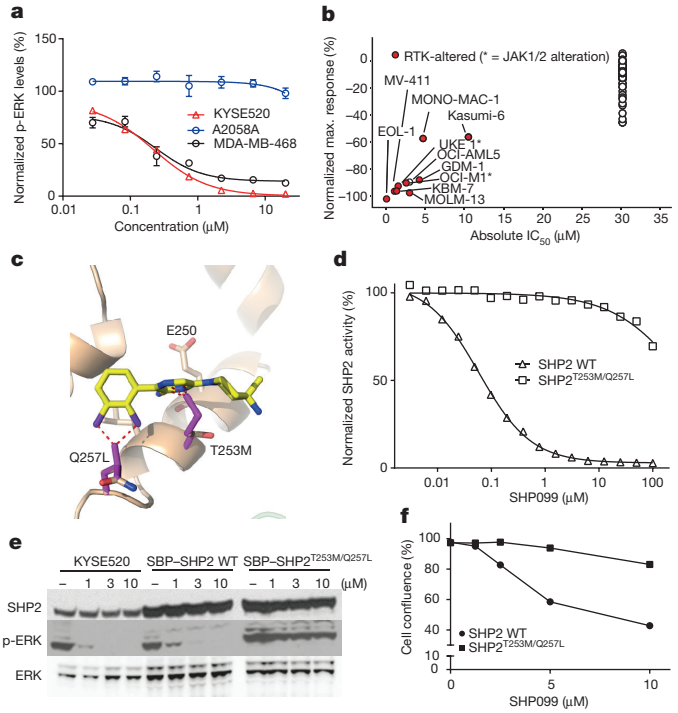


Figure 3 | Validation of SHP2-dependent inhibitory activity of SHP009 in cells. a, Inhibition of p-ERK activity by SHP099 in A2058, KYSE520 or MDA-MB-468 cells assayed by SureFire p-ERK assay. p-ERK activity is expressed as a percentage of the DMSO control. Data presented as mean ± s.d. ($n = 3$). **b**, Activity of SHP099 in 71 haematopoietic cell lines. The data are plotted as normalized inhibition at 30 μM SHP099 (y axis) against calculated absolute IC_{50} values of SHP099 for each cell lines (x axis). Solid red circles represent cancer cells with RTK, JAK1 or JAK2 mutations, black circles correspond to *NRAS*- or *KRAS*-mutated cells. The corresponding data and cell line genotypes are in Supplementary Information Table 1. **c**, Model of engineered $SHP2^{T253M/Q257L}$ mutant highlighting steric clashes between the mutated residues and SHP099. **d**, Biochemical inhibition of wild-type SHP2 and $SHP2^{T253M/Q257L}$ by SHP099. **e**, Western blot of SHP2, p-ERK and ERK from SHP2-depleted KYSE520 cells stably re-expressing SBP-tagged wild-type SHP2 or SBP–$SHP2^{T253M/Q257L}$ and treated with SHP099 (1, 3, 10 μM). **f**, Proliferation of SHP2-depleted KYSE520 cells stably re-expressing wild-type streptavidin-binding peptide tagged (SBP) SHP2 or SBP–$SHP2^{T253M/Q257L}$ treated with SHP099 (1.25, 2.5, 5, 10 μM). Data points along the line in **d** and **f** represent the mean of two replicate values.

expected cell growth inhibition (Extended Data Fig. 4c). We further profiled SHP099 in a panel of 71 haematopoietic cancer cell lines and 26 colorectal cancer cell lines. Haematopoietic cancer cells with known alterations in oncogenic RTKs or other tyrosine kinases such as JAK1 or JAK2 were sensitive to SHP099 inhibition (Fig. 3b, Supplementary Information Table 1). Similarly, colorectal cancer cells that were sensitive to the potent Her1/2 and EGFR inhibitor Lapatinib, and hence dependent on EGFR signalling, also responded to SHP099 treatment (Extended Data Fig. 4d, Supplementary Information Table 2). In contrast, *RAS*- or *BRAF*-mutated cells from both lineages were generally resistant to SHP099 treatment ($IC_{50} > 30\,\mu M$ for haematopoietic lines and >20 μM for colorectal lines) (Fig. 3b, Extended Data Fig. 4d and Supplementary Table 2). The observed correlation between RTK dependence and SHP099 sensitivity is robustly supported by a Chi-squared test of independence ($P < 1.1 \times 10^{-15}$). These data therefore recapitulate the differential growth inhibitory effects observed in the shRNA screen and suggest a strong association between RTK dependence and sensitivity to SHP2 inhibition.

To determine whether SHP099-mediated suppression of MAPK signalling and growth inhibition was an on-target consequence of

© 2016 Macmillan Publishers Limited. All rights reserved

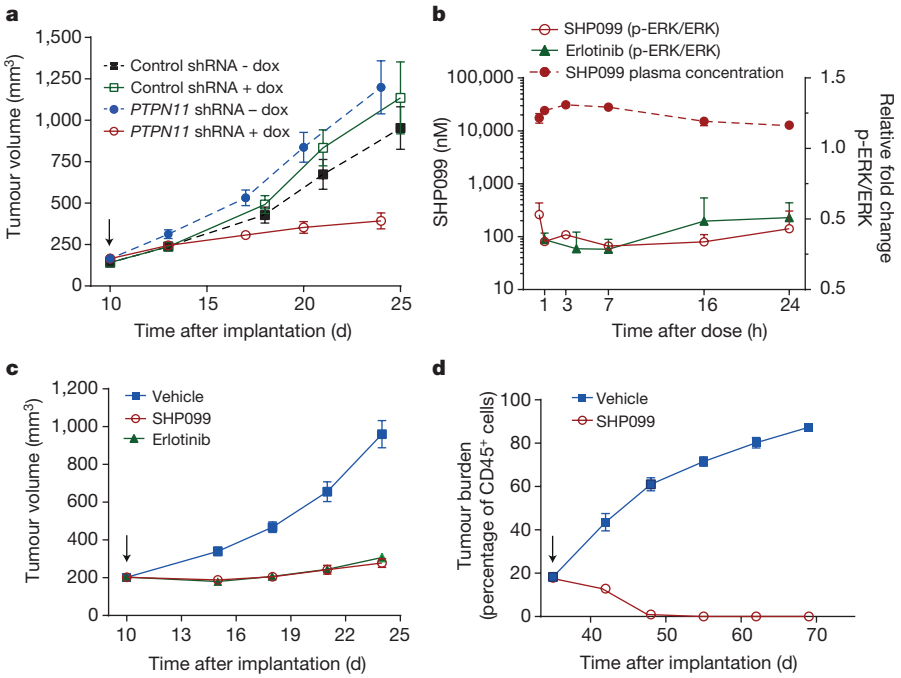


Figure 4 | *In vitro* and *in vivo* characterization of SHP099. a, SHP2 knockdown inhibits the growth of established KYSE520 xenograft tumours *in vivo*. KYSE520 cells stably expressing dox-inducible non-targeting control or *PTPN11* shRNA were inoculated into mice. Mice were treated with vehicle (dotted lines) or dox-supplemented diet (solid lines) starting 10 days after implantation. The tumour volume of vehicle or dox-treated mice is plotted as the mean ± s.e.m. ($n = 9$). **b**, *In vivo* plasma SHP099 concentration and xenograft p-ERK levels following a single oral administration of SHP099 (100 mg per kg) or erlotinib (80 mg per kg) to nude mice with subcutaneous KYSE520 xenografts. SHP099 plasma concentrations and xenograft p-ERK level were assessed through the first 24 h following compound administration. Data are plotted as the mean ± s.e.m. ($n = 3$). **c**, Antitumour efficacy of SHP099 (100 mg per kg daily) or erlotinib (80 mg per kg daily) in nude mice bearing established subcutaneous KYSE520 xenografts. Mice were administered compounds daily by oral gavage starting 10 days after cell implantation. Data are plotted as the group mean ± s.e.m. ($n = 9$). **d**, Antitumour efficacy SHP099 (75 mg per kg) in immunocompromised mice with an orthotopic, primary-tumour-derived AML xenograft. Mice were administered SHP099 daily by oral gavage starting 35 days after tumour implantation and continued for 34 days. Tumour burden was assessed by FACS detection of $hCD45^+$ leukaemic cells in circulation. Data are plotted as the group mean ± s.e.m. ($n = 7$). The arrow in **a**, **c** and **d** denotes the start of SHP099 treatment.

pharmacological inhibition of SHP2, inhibitor-resistant alleles were developed. From the co-crystal structure, we hypothesized that mutation of Thr253 and Gln257 would disrupt SHP099 binding, but maintain the integrity of the three-domain regulatory interface (Fig. 3c). After testing single- and double-mutant alleles, $SHP2^{T253M/Q257L}$ was found to retain the catalytic activity and auto-inhibited basal state of SHP2, but was 1,000-fold less sensitive to SHP099 inhibition as compared to wild-type SHP2 *in vitro* (Fig. 3d, Extended Data Fig. 3d). Treatment of KYSE520 cells expressing $SHP2^{T253M/Q257L}$ with SHP099 failed to inhibit both the signalling to p-ERK and cellular growth compared to the KYSE520 wild-type SHP2 control (Fig. 3e, f). These data strongly suggest that SHP099 inhibits MAPK signalling and proliferation in RTK-dependent cells through direct on-target inhibition of SHP2.

We next established a subcutaneous xenograft model using KYSE520 stably transduced with dox-inducible *PTPN11* shRNA to investigate whether SHP2 was required for tumour maintenance *in vivo*. Expression of *PTPN11* shRNA was induced by dox when the tumour volume reached ~200 mm^3. SHP2 knockdown led to a significant reduction in p-ERK levels and marked tumour growth inhibition ($P < 0.05$), whereas a control non-targeting shRNA showed no effect (Fig. 4a and Extended Data Fig. 5a, b).

On the basis of its potent effects in cell culture, we next evaluated the efficacy of SHP099 in nude mice with established, subcutaneous KYSE520 xenografts. Following a single 100 mg per kg (body weight) oral dose, SHP099 yielded free plasma concentrations >10 μM and was associated with a robust inhibition of p-ERK (>50%) that was maintained for 24 h after the dose was administered (Fig. 4b). At this exposure, SHP099 was predicted to achieve a significant anti-proliferative effect on the basis of *in vitro* characterization (Fig. 3a, b and Extended Data Fig. 4a, b). SHP099 was administered by oral gavage at 100 mg per kg daily to nude mice with KYSE520 xenografts and yielded marked tumour growth inhibition (Fig. 4c) over a 24-day time period. In a follow-up study, orally administered SHP099 showed dose-dependent anti-tumour activity in the KYSE520 xenograft model and was well tolerated, as demonstrated by insignificant or no body weight loss over the entire course of treatment (Extended Data Fig. 5c, d). For comparison, we treated a parallel cohort of KYSE520 tumour-bearing mice with the EGFR inhibitor erlotinib. The p-ERK modulation and tumour growth inhibition observed with erlotinib was equivalent to that observed with SHP099 (Fig. 4b, c). To extend this observation to the setting of RTK activation in haematological malignancies, SHP099 was evaluated in an orthotopic human-primary-tumour-derived FLT3-ITD acute myeloid leukaemia (AML) model. Here, daily dosing at 75 mg per kg led to near-complete eradication of circulating human (h)$CD45^+$ leukaemic cells (Fig. 4d) and significantly reduced splenomegaly in the mice (Extended Data Fig. 5e, f). In summary, pharmacological inhibition of SHP2 by SHP099 is efficacious and well tolerated and therefore offers a novel therapeutic approach to target RTK-dependent cancers.

Despite two decades of research describing the central role of SHP2 in developmental and oncogenic signalling pathways, no SHP2 inhibitor has progressed to clinical use. Although catalytic SHP2 inhibitors have been described[11–15], they are typically of low potency and inhibit other phosphatases. Although the allosteric inhibition of a metal-dependent serine/threonine phosphatase family has been explored[21], SHP099 is the first example of a potent, selective and orally bioavailable allosteric PTP inhibitor specific to SHP2 that is efficacious and well tolerated in patient-derived tumour xenograft models. Our study provides evidence that pharmacological inhibition of SHP2 is a viable

© 2016 Macmillan Publishers Limited. All rights reserved

strategy to target RTK-driven cancers and presents a new chemical tool for further interrogation of the multifaceted cellular functions of SHP2 in development, tumorigenesis, RTK-driven drug resistance and immune-checkpoint modulation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 November 2015; accepted 26 May 2016.
Published online 29 June 2016.

1. Grossmann, K. S., Rosário, M., Birchmeier, C. & Birchmeier, W. The tyrosine phosphatase Shp2 in development and cancer. *Adv. Cancer Res.* **106,** 53–89 (2010).
2. Chan, R. J. & Feng, G. S. *PTPN11* is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood* **109,** 862–867 (2007).
3. Matozaki, T., Murata, Y., Saito, Y., Okazawa, H. & Ohnishi, H. Protein tyrosine phosphatase SHP-2: a proto-oncogene product that promotes Ras activation. *Cancer Sci.* **100,** 1786–1793 (2009).
4. Mohi, M. G. & Neel, B. G. The role of Shp2 (*PTPN11*) in cancer. *Curr. Opin. Genet. Dev.* **17,** 23–30 (2007).
5. Östman, A., Hellberg, C. & Böhmer, F. D. Protein-tyrosine phosphatases and cancer. *Nat. Rev. Cancer* **6,** 307–320 (2006).
6. Gavrieli, M., Watanabe, N., Loftin, S. K., Murphy, T. L. & Murphy, K. M. Characterization of phosphotyrosine binding motifs in the cytoplasmic domain of B and T lymphocyte attenuator required for association with protein tyrosine phosphatases SHP-1 and SHP-2. *Biochem. Biophys. Res. Commun.* **312,** 1236–1243 (2003).
7. Okazaki, T., Chikuma, S., Iwai, Y., Fagarasan, S. & Honjo, T. A rheostat for immune responses: the unique properties of PD-1 and their advantages for clinical application. *Nat. Immunol.* **14,** 1212–1218 (2013).
8. Prahallad, A. *et al.* PTPN11 is a central node in intrinsic and acquired resistance to targeted cancer drugs. *Cell Reports* **12,** 1978–1985 (2015).
9. Schneeberger, V. E. *et al.* Inhibition of Shp2 suppresses mutant EGFR-induced lung tumors in transgenic mouse model of lung adenocarcinoma. *Oncotarget* **6,** 6191–6202 (2015).
10. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483,** 603–607 (2012).
11. Scott, L. M. *et al.* Shp2 protein tyrosine phosphatase inhibitor activity of estramustine phosphate and its triterpenoid analogs. *Bioorg. Med. Chem. Lett.* **21,** 730–733 (2011).
12. Grosskopf, S. *et al.* Selective inhibitors of the protein tyrosine phosphatase SHP2 block cellular motility and growth of cancer cells *in vitro* and *in vivo*. *ChemMedChem* **10,** 815–826 (2015).
13. He, R. *et al.* Exploring the existing drug space for novel pTyr mimetic and SHP2 inhibitors. *ACS Med. Chem. Lett.* **6,** 782–786 (2015).
14. Hellmuth, K. *et al.* Specific inhibitors of the protein tyrosine phosphatase Shp2 identified by high-throughput docking. *Proc. Natl Acad. Sci. USA* **105,** 7275–7280 (2008).
15. Zeng, L. F. *et al.* Therapeutic potential of targeting the oncogenic SHP2 phosphatase. *J. Med. Chem.* **57,** 6594–6609 (2014).
16. Pluskey, S., Wandless, T. J., Walsh, C. T. & Shoelson, S. E. Potent stimulation of SH-PTP2 phosphatase activity by simultaneous occupancy of both SH2 domains. *J. Biol. Chem.* **270,** 2897–2900 (1995).
17. Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M. J. & Shoelson, S. E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* **92,** 441–450 (1998).
18. Manley, P. W. *et al.* Extended kinase profile and properties of the protein kinase inhibitor nilotinib. *Biochim. Biophys. Acta.* **1804,** 445–453 (2010).
19. Bender, A. *et al.* Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* **2,** 861–873 (2007).
20. Szczepankiewicz, B. G. *et al.* Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* **125,** 4087–4096 (2003).
21. Gilmartin, A. G. *et al.* Allosteric Wip1 phosphatase inhibition through flap-subdomain interaction. *Nat. Chem. Biol.* **10,** 181–187 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements Use of the IMCA-CAT beamline 17-ID at the Advanced Photon Source was supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Hauptman-Woodward Medical Research Institute. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number DE-AC02-06CH11357.

Author Contributions Y.P.C., F.F., H.-X.H., K.H., S.L., J.S., P.Z., H.M.C. performed or directed cellular assays data generation and analysis; P.F., M.G.A., Z.B.K., S.H., E.P., C.Q., S.S, P.W., J.-H.Z. and P.D.F. performed or directed biochemical experiments; B.A., V.G.C., B.F., H.G., L.R.L.B., M.J.M., M.P., G.Y. and J.Y. performed or directed *in vivo* pharmacology or pharmacokinetic/pharmacodynamics experiments and data analysis; J.R.D. and K.V. directed or performed bioinformatics analyses; M.J.L., J.G.-F., C.F., C.H.-T.C, Z.C., D.G., R.K., M.K., J.L., F.L., G.L., D.M., M.P., L.P., M.D.S., T.S., S.W. Designed, synthesized and/or directed the design or synthesis of SHP2 inhibitors; Z.D., M.K., and S.W. performed protein and inhibitor structural modelling or cheminformatics analyses; M.F., J.J. and T.S. designed, directed or performed biophysics experiments; M.F. and T.S. directed or performed x-ray crystallography experiments; Y.P.C., J.R.D., L.R.L.B., M.F., M.J.M., K.V., H.M.C., T.S., W.R.S. and P.D.F. prepared figures and tables for the main text and Supplementary Information; Y.P.C., M.J.L., J.R.D., L.R.L.B., M.J.M., K.V., N.J.K., H.M.C., T.S., W.R.S. and P.D.F. wrote and edited the main text and Supplementary Information; P.D.F., N.J.K., T.R., T.S., W.R.S. and M.W. contributed to overall project oversight.

Author Information Atomic coordinates and structure factors for the SHP2–SHP099 binary complex structure have been deposited with the Protein Data Bank under accession number 5EHR. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.D.F. (pafortin@gmail.com), T.S. (travis.stams@novartis.com) or W.R.S. (william.sellers@novartis.com) .

Reviewer Information *Nature* thanks B. Neel and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during outcome assessment.

Bioinformatics. Statistical analyses were performed as follows: cell counts from the pooled shRNA experiments were normalized using a quantile normalization procedure as described elsewhere[22] and normalized scores for shRNA targeting the same gene were aggregated at the gene level using the ATARIS algorithm[23]. ATARIS scores for genes of interest were binned into three categories, 'dependent', 'independent' and 'unclear' on the basis of the degree of dropout. Most RTKs show a strong phenotype in these shRNA assays, here cell lines with ATARIS score <-1 for any of the RTK genes (*EGFR*, *FRS2*, *cMET*, *ERBB2*, *FLT3*) were considered RTK-dependent, cell lines with ATARIS score >0 were considered to be 'RTK-independent' and cell lines with ATARIS scores in between 0 and -1 were considered 'unclear' and removed from statistical analyses to remove weaker effects. An identical approach was used for comparing with RAS/RAF (BRAF, NRAS, and KRAS). SHP2 shows a less marked shRNA phenotype in these shRNA assays and the ATARIS score threshold for assigning SHP2-dependence was set to <-0.8.

Association analyses were done using a Fisher's exact test to assess (1) the association of SHP2-dependence with RTK-dependence and (2) the association of SHP2-dependence with RTK-alterations, and the odds ratio and *P* value were reported. The statistical analysis for the association of RTK-dependence and SHP099 sensitivity was performed as follows. In cell lines derived from haematological cancers, we defined RTK-dependence as RTK-altered/driven growth and in CRC lines as lapatanib sensitivity ($IC_{50} \leq 0.5\,\mu M$). The *P* value was derived using the following relationships observed between RTK-dependence and sensitivity to SHP099 ($IC_{50} \leq 10\,\mu M$) in all 97 lines: 80 lines are SHP099 insensitive (69 expected based on RTK independence and RAS/RAF mutation status), 0 lines are RTK-dependent and not SHP099 sensitive (10 expected), 4 lines are RTK-independent and SHP099 sensitive (14 expected), and 13 lines are both RTK-dependent and SHP099 sensitive (2 expected).

Cell culture, viral production and infection. Human cancer cell lines originated from the CCLE, authenticated by single-nucleotide polymorphism analysis and tested for mycoplasma infection[10]. Sum52, KYSE520, MDA-MB-468, KATO III, NCI-H2170, NCI-H2170, NCI-H2228, MDA-MB-231 and A2058 were cultured in RPMI medium (Invitrogen) supplemented with 10% fetal bovine serum (Invitrogen). H293 cells were grown in DMEM medium (Invitrogen) supplemented with 10% fetal bovine serum. For viral production, vectors pshSHP2 or pshNTC (nontargeting control) were transfected in H293 cells using TransIT-293 transfection reagent (Mirus Inc.), following the manufacturers protocol. At 72 h after transfection, the cell culture medium was filtered through a 0.45 μm filter, and the viral supernatant supplemented with $8\,\mu g\,ml^{-1}$ of polybrene was used for the infection of cells. For viral infection, ~70% confluent cells in six-well dishes were infected with virus at a multiplicity of infection of 5 units per cell for 4 h and allowed to recover for 24 h with fresh medium. Stable clones were selected using either puromycin or G418. Methods for profiling small molecule inhibitors in the haematopoietic cells are described elsewhere[10]. Cells were treated using SHP099 concentration varying from 0 to 30 μM. Cellular viability was measured using CellTiter-Glo.

SHP2 allosteric inhibition assay. SHP2 is allosterically activated through binding of bistyrosylphorphorylated peptides to its Src Homology 2 (SH2) domains. The latter activation step leads to the release of the auto-inhibitory interface of SHP2, which in turn renders the SHP2 protein tyrosine phosphatase (PTP) active and available for substrate recognition and reaction catalysis. The catalytic activity of SHP2 was monitored using the surrogate substrate DiFMUP in a prompt fluorescence assay format. More specifically, the phosphatase reactions were performed at room temperature in 384-well black polystyrene plate, flat bottom, low flange, non-binding surface (Corning) using a final reaction volume of 25 μl and the following assay buffer conditions: 60 mM HEPES, pH 7.2, 75 mM NaCl, 75 mM KCl, 1 mM EDTA, 0.05% P-20, 5 mM DTT. 0.5 nM of SHP2 was co-incubated with of 0.5 μM of bisphosphorylated IRS1 peptide (sequence: H2N-LN(pY)IDLDLV(dPEG8)LST(pY)ASINFQK-amide) and 0.003–100 μM of the inhibitory compounds. After 30–60 min incubation at 25 °C, the surrogate substrate DiFMUP (Invitrogen) was added to the reaction and incubated at 25 °C for 30 min. The reaction was then quenched by the addition of 5 μl of a 160 μM solution of bpV(Phen) (Enzo Life Sciences). The fluorescence signal was monitored using a microplate reader (Envision, Perkin-Elmer) using excitation and emission wavelengths of 340 nm and 450 nm, respectively. The inhibitor dose–response curves were analysed using normalized IC_{50} regression curve fitting with control-based normalization.

Protein expression and purification. Two constructs of human SHP2 (accession number NP_002825.3) were generated by cloning the *PTPN11* gene encoding truncations Met1–Leu525 (named SHP2) and Ala237–Ile529 (named SHP^{PTP}) into a pET30 vector. A coding sequence for a 6× histidine tag followed by a tomato etch virus (TEV) protease consensus sequence was added 5′ to the constructs sequence. The construct was transformed into BL21 Star (DE3) cells and grown at 37 °C in Terrific Broth containing $100\,\mu g\,ml^{-1}$ kanamycin. At an OD_{600} of 4.0, SHP2 expression was induced using 1 mM IPTG. Cells were collected after overnight growth at 18 °C.

Cell pellets were resuspended in lysis buffer containing 50 mM Tris-HCl (pH 8.5), 25 mM imidazole, 500 mM NaCl, 2.5 mM $MgCl_2$, 1 mM TCEP, $1\,\mu g\,ml^{-1}$ DNase1, and complete EDTA-free protease inhibitor and lysed using a microfluidizer, followed by ultracentrifugation. The supernatant was loaded onto a HisTrap HP chelating column in 50 mM Tris-HCl, 25 mM imidazole, 500 mM NaCl, 1 mM TCEP and protein was eluted with the addition of 250 mM imidazole. The N-terminal histidine tag was removed with an overnight incubation using TEV protease at 4 °C. The protein was subsequently diluted to 50 mM NaCl with 20 mM Tris-HCl (pH 8.5), 1 mM TCEP then applied to a HiTrap Q FastFlow column equilibrated with 20 mM Tris (pH 8.5), 50 mM NaCl, 1 mM TCEP. The protein was eluted with a 10-column volume gradient from 50–500 mM NaCl. Fractions containing SHP2 were pooled and concentrated then loaded onto a HiLoad Superdex200 PG 16/100 column, exchanging the protein into the crystallization buffer, 20 mM Tris-HCl (pH 8.5), 150 mM NaCl and 3 mM TCEP. The protein was concentrated to $15\,mg\,ml^{-1}$ for use in crystallization. QuickChange mutagenesis (Agilent) was used to generate the $SHP2^{T253M/Q257L}$ mutant using the above construct, and same procedure for expression and purification.

Chemistry. All solvents employed were commercially available 'anhydrous' grade, and reagents were used as received unless otherwise noted. A Biotage Initiator Sixty system was used for microwave heating. Flash column chromatography was performed on either an Analogix Intelliflash 280 using Si 50 columns (32–63 μm, 230–400 mesh, 60 Å) or on a Biotage SP1 system (32–63 μm particle size, KP-Sil, 60 Å pore size). Preparative high pressure liquid chromatography (HPLC) was performed using a Waters 2525 pump with 2487 dual wavelength detector and 2767 sample manager. Columns were Waters C18 OBD 5 μm, either 50 × 100 mm Xbridge or 30 × 100 mm Sunfire. NMR spectra were recorded on a Bruker AV400 (Avance 400 MHz) or AV600 (Avance 600 MHz) instruments. Analytical LC-MS was conducted using an Agilent 1100 series with UV detection at 214 nm and 254 nm, and an electrospray mode (ESI) coupled with a Waters ZQ single quad mass detector. One of two methods was used: (1) 3 μl of sample was injected on an inertsil C8 3 cm × 5 mm × 3 μm and eluted using a 5:95 to 95:5 acetonitrile:H_2O (5 mM ammonium formate) 2 min gradient; (2) 3 μl of sample was injected on an inertsil C8 3 cm × 5 mm × 3 μm and eluted using a 20:80 to5:95 acetonitrile:H_2O (10 mM ammonium formate) 2 min gradient. The purity of all tested compounds was determined by LC/ESI-MS data recorded using an Agilent 6220 mass spectrometer with electrospray ionization source and Agilent 1200 liquid chromatography. The mass accuracy of the system has been found to be <5 ppm. HPLC separation was performed at $75\,ml\,min^{-1}$ flow rate with the indicated gradient within 3.5 min with an initial hold of 10 s. 10 mM ammonia hydroxide or 0.1 M TFA was used as the modifier additive in the aqueous phase. All compounds were found to be >95% purity.

SHP836 originated from a purchased chemical library included in the general Novartis screening chemical library. SHP836 is also known as GW286103 (ref. 24). SHP043 was synthesized as previously described[25].

6-(4-amino-4-methylpiperidin-1-yl)-3-(2,3-dichlorophenyl)pyrazin-2-amine (SHP099). A mixture of 3-bromo-6-chloropyrazine-2-amine (1.5 g, 7.2 mmol), (2,3-dichlorophenyl)boronic acid (1.37 g, 7.2 mmol), PdCl2(dppf).DCM adduct (294 mg, 0.36 mmol), and potassium phosphate (4.58 g, 21.59 mmol) in MeCN:H_2O (9:1, 15 ml) was stirred for 4 h at 120 °C. After cooling to room temperature, the reaction mixture was filtered through a pad of Celite followed by EtOAc wash. The solvent was removed under reduced pressure and the resulting residue was purified by silica chromatography (5 to 30% gradient of EtOAc in heptane) to give 6-chloro-3-(2,3-dichlorophenyl)pyrazin-2-amine (1.46 g, 5.32 mmol) as yellow solid. ^1H NMR (400 MHz, chloroform-*d*) δ ppm 7.99–8.08 (m, 1H), 7.62 (dd, $J = 7.78, 1.76$ Hz, 1 H), 7.36–7.42 (m, 1H), 7.32–7.36 (m, 1H), 4.69 (br s, 2H). ^{13}C NMR (101 MHz, chloroform-*d*) δ ppm 151.58, 146.66, 136.56, 136.30, 134.24, 132.13, 131.78, 131.56, 129.34, 128.32. HRMS calculated for $C_{10}H_7Cl_3N_3$ $(M+H)^+$ 273.9706, found 273.9706.

A mixture of 6-chloro-3-(2,3-dichlorophenyl)pyrazin-2-amine (125 mg, 0.455 mmol), *tert*-butyl (4-methylpiperidin-4-yl)carbamate (195 mg, 0.911 mmol), and potassium phosphate (97 mg, 0.455 mmol) in NMP (1 ml) was stirred for 36 h at 140 °C. After cooling to room temperature, the mixture was poured into a separation funnel containing saturated aqueous NH_4Cl and it was extracted with EtOAc (3 × 5 ml). The combined organic phases were dried over $MgSO_4$, filtered and the solvents were removed under reduced pressure. The

© 2016 Macmillan Publishers Limited. All rights reserved

resulting residue was by silica chromatography (5 to 30% gradient of EtOAc in heptane) to give *tert*-butyl (1-(6-amino-5-(2,3-dichlorophenyl)pyrazin-2-yl)-4-methylpiperidin-4-yl)carbamate (113 mg, 0.250 mmol) as yellow solid. ^{1}H NMR (400 MHz, chloroform-*d*) δ ppm 7.54 (s, 1H), 7.43 (dd, $J=7.78$, 2.01 Hz, 1H), 7.16–7.29 (m, 2H), 4.36 (br s, 1H), 4.17 (s, 2H), 3.80 (dt, $J=13.68$, 4.33 Hz, 2H), 3.18–3.31 (m, 2H), 2.03 (br d, $J=13.30$ Hz, 2H), 1.59 (ddd, $J=13.87$, 10.23, 4.27 Hz, 2H), 1.35–1.42 (m, 9H), 1.32–1.35 (m, 3H). ^{13}C NMR (101 MHz, chloroform-*d*) δ ppm 154.47, 153.58, 149.85, 138.65, 133.72, 132.40, 130.33, 130.11, 127.89, 125.61, 119.19, 79.23, 50.57, 40.60 (×2), 35.83 (×2), 28.46 (×3), 28.21. HRMS calculated for $C_{21}H_{28}Cl_2N_5O_2$ $(M+H)^+$ 454.1591, found 454.1591.

A solution of *tert*-butyl (1-(6-amino-5-(2,3-dichlorophenyl)pyrazin-2-yl)-4-methylpiperidin-4-yl)carbamate (113 mg, 0.250 mmol) in THF:H_2O (4:1, 2.5 ml) was treated with HCl (4 M in dioxane, 230 µl, 0.928 mmol). The resulting mixture was stirred for 2 h at 140 °C. After cooling to room temperature, the volatiles were removed under reduced pressure, and the resulting residue was diluted with EtOAc (10 ml), H_2O (10 ml). The phases were separated and the aqueous was further extracted with EtOAc (2 × 5 ml). The combined organic phases were discarded; the aqueous phase was basified to pH 9 with NaOH 2 M, and extracted with EtOAc (3 × 10 mL). The combined organic phases were dried over $MgSO_4$, filtered and the solvent was removed under reduced pressure to afford 6-(4-amino-4-methylpiperidin-1-yl)-3-(2,3-dichlorophenyl)pyrazin-2-amine (84 mg, 0.238 mmol) as a yellow solid. ^{1}H NMR (400 MHz, methanol-d_4) δ ppm 7.61 (dd, $J=7.91$, 1.63 Hz, 1H) 7.47 (s, 1H) 7.40 (t, $J=7.78$ Hz, 1H) 7.34 (dd, $J=7.65$, 1.63 Hz, 1H) 3.78 (ddd, $J=13.43$, 7.15, 4.27 Hz, 2H) 3.50–3.64 (m, 2H) 1.55–1.75 (m, 4H) 1.25 (s, 3 H). ^{13}C NMR (101 MHz, methanol-d_4) δ ppm 155.42, 152.89, 139.87, 134.55, 133.90, 131.60, 131.48, 129.24, 125.78, 117.54, 54.84, 42.18 (×2), 39.40 (×2), 27.55. HRMS calculated for $C_{16}H_{20}Cl_2N$ $(M+H)^+$ 352.1096, found 352.1099.

Crystallization and structure determination. Sitting drop vapour diffusion method was used for crystallization, with the crystallization well containing 17% PEG 3350 and 200 mM ammonium phosphate and a drop with a 1:1 volume of SHP2 protein and crystallization solution. Crystals were formed within five days, and subsequently soaked in the crystallization solution with 2.5 mM SHP099. This was followed by cryoprotection using the crystallization solution with the addition of 20% glycerol and 1 mM of compound 1, followed by flash freezing directly into liquid nitrogen.

Diffraction data for the SHP2–SHP099 complex were collected on a Dectris Pilatus 6M Detector at beamline 17ID (IMCA-CAT) at the Advanced Photon Source at Argonne National Laboratories. The data were measured from a single crystal maintained at 100 K at a wavelength of 1 Å, and the reflections were indexed, integrated, and scaled using XDS[26]. The space group of the complex was $P2_1$ with two molecules in the asymmetric unit. The structure was determined with Fourier methods, using the SHP2 apo structure[17] (PDB accession 2SHP) with all waters removed. Structure determination was achieved through iterative rounds of positional refinement and model building using BUSTER[27] and COOT[28], yielding the published SHP2–SHP099 binary complex structure (PDB accession number 5EHR). Individual B-factors were refined using an overall anisotropic B-factor refinement along with bulk solvent correction. The solvent, phosphate ions, and inhibitor were built into the density in later rounds of the refinement. Data collection and refinement statistics are shown in Extended Data Table 4.

Isothermal titration calorimetry. The binding of SHP099 was studied by isothermal titration calorimetry (ITC) using the Auto iTC-200 calorimeter from Malvern Instruments. SHP2 was dialysed and SHP099 was dissolved into the identical buffer composed of 25 mM Hepes (pH 7.5), 100 mM NaCl, and 0.25 mM TCEP with 2% DMSO. The titration was performed at 25 °C by injecting 2.5 µl aliquots of SHP099 into the calorimetric cell (~200 µl) containing the protein at a concentration of 55 µM. The concentration of SHP099 in the syringe was 450 µM. The heat evolved upon each injection was obtained from the integral of the calorimetric signal. The individual heats were plotted against the molar ratio, and the enthalpy change (ΔH) and association constant ($K_a = 1/K_d$) were obtained by nonlinear regression of the data.

p-ERK cellular assay. p-ERK cellular assay was using the AlphaScreen *SureFire* Phospho-ERK 1/2 Kit (PerkinElmer): A2058, KYSE520 or MDA-MB-468 cells (30,000 cells per well) were grown in 96-well plate culture overnight and treated with SHP099 at concentrations of 20, 6.6, 2.2, 0.74, 0.24,0.08, 0.027 µM for 2 h at 37 °C. Incubations were terminated by addition of 30 µl of lysis buffer (PerkinElmer) supplied with the SureFire phospho-extracellular signal-regulated kinase (p-ERK) assay kit (PerkinElmer). Samples were processed according to the manufacturer's directions. The fluorescence signal from pERK was measured in duplicate using a 2101 multi-label reader (PerkinElmer EnVision). The percentage of inhibition was normalized by the total ERK signal and compared with the DMSO vehicle control.

Colony formation assay and cell proliferation assay. KYSE520, MDA-MB-468, A2058, Sum52, KatoIII cells (1500 cells per well) were plated onto 24-well plates in 300 µl medium (RPMI-1640 containing 10% FBS, Lonza). For drug treatment, SHP099 were added at various concentrations (10, 5, 2.5, 1.25 µM) 24 h and 5 days after cell plating. At day 11, colonies were stained with 0.2% crystal violet (MP Biomedicals) and subsequently dissolved in 20% acetic acid for quantitation using a Spectramax reader (Thermo Scientific). In cell proliferation assay, KYSE520, A2058 and colorectal cancer cells (1500 cells per well) were plated onto 96-well plates in 100 µl medium (RPMI-1640 containing 10% FBS, Lonza) and treated with SHP099 and/or lapatinib concentration varying from 0.0 to 20.25 µM. At day 5, 50 µl CellTiter-Glo reagent (Promega) was added, and the luminescent signal was determined according to the supplier's instruction (Promega). Method of profiling SHP099 in the haematopoietic cancer cell panels were described previously[10]. Cells were treated using SHP099 concentration varying from 0.0 to 30 µM. Cellular viability was measured using CellTiter-Glo at day 3.

Western blotting. Cells were lysed on ice for 30 min with CST lysis buffer (Cell Signaling) containing phosSTOP (Roche). Cell lysates were centrifuged at 4 °C for 15 min with a microfuge. Protein concentrations of cell lysate supernatants were measured. Cell lysate supernatants of equal-amount proteins were used for immunoblotting. The following antibodies were used: SHP2 (Santa Cruz SC-280), pERK (CST #4377S), ERK (Santa Cruz SC-93), RAS (CST#3965S), pAKT (CST #4060S), GAPDH (CST #2118S).

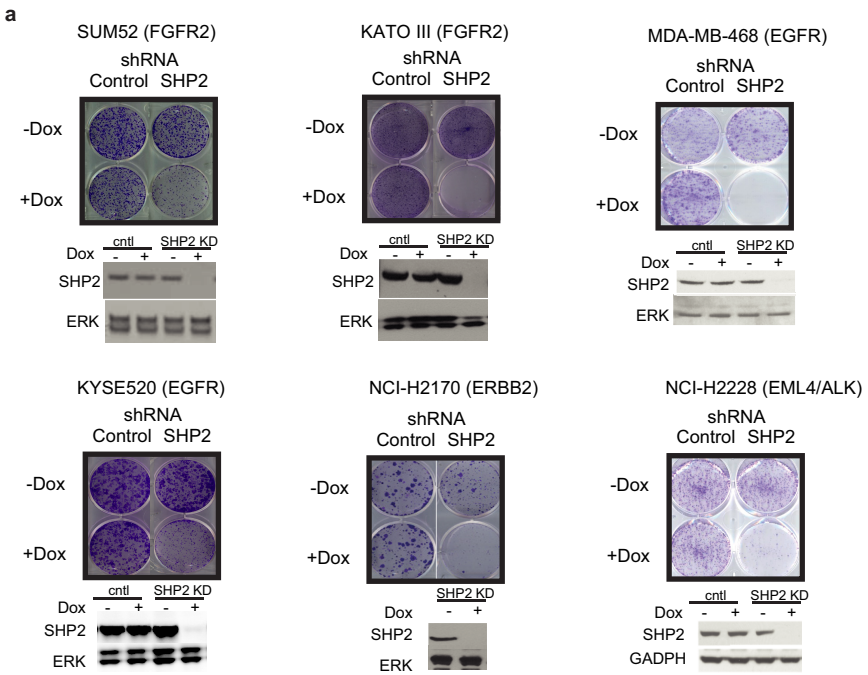
Tumour xenograft experiments and tissue-sample preparations. All animal studies were carried out according to the Novartis Guide for the Care and Use of Laboratory Animals. Cell lines were confirmed to be devoid of mycoplasma and mouse viruses before use. Sample sizes were determined roughly on the basis of a power analysis using historical internal xenograft tumour volume data and anti-tumour responses. In the efficacy studies, animals were randomly assigned to treatment groups by an algorithm that moves animals around to achieve the best case distribution to assure that each treatment group has similar mean tumour burden and standard deviation. Female athymic nude mice (9–12 weeks of age) were inoculated subcutaneously (3×10^6 cells in a suspension containing 50% phenol red-free matrigel (BD Biosciences) in Hank's balanced salt solution) with parental KYSE520 cells or KYSE520 cells stably expressing dox-inducible control non-targeting shRNA or distinct *PTPN11*-targeting shRNA. For pharmacokinetic/pharmacodynamics studies, mice were administered a single dose of vehicle control, erlotinib, or SHP099 by oral gavage once tumours reached roughly 500 mm^3. Mice were subsequently killed at predetermined time points following a single dose of compound, at which point plasma and xenograft fragments were collected for determination of SHP099 concentrations and p-ERK modulation, respectively. For efficacy studies, xenograft tumours were measured twice weekly by calipering in two dimensions. Once tumours reached roughly 200 mm^3, mice were randomly assigned to treatment groups. For the shRNA study, on day 10 after cell line implantation, mice were assigned to receive either vehicle diet (standard diet) or dox-supplemented diet (Mod LabDiet 5053, 400 p.p.m. dox) for the duration of the study. For the efficacy study, on day 10 after cell line implantation, mice were assigned to receive either vehicle, SHP099 (100 mg per kg daily), or erlotinib (80 mg per kg daily) by oral gavage. In both efficacy studies, tumour volume and mouse body weight was assessed twice weekly. To assess MAPK pathway modulation in xenograft protein lysates, total and phospho ERK1/2 was assessed using a commercially available kit (Meso Scale Discovery catalogue number K15107D). The assay was conducted as recommended by Meso Scale Discovery with the exception that protein lysate was incubated overnight. The development of the patient-derived AML xenograft model in mice and study design has been previously described[29]. An additional group of mice ($n=7$) was added to the study on day 35 after tumour implantation and treated with SHP099 (75 mg per kg daily) for 34 days. At the end of the study (69 days after tumour implantation), mice were euthanized, and spleen weights of individual mice were recorded. In all cases, no data or animals were excluded and results are expressed as mean and standard error of the mean. No further statistical analysis was performed.

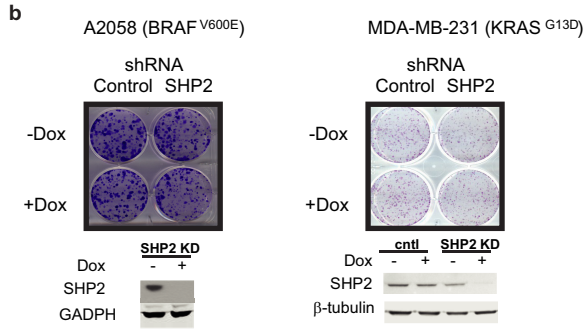
Pharmacokinetics. Plasma samples were precipitated and diluted with acetonitrile containing internal standard and prepared for LC-MS/MS. An aliquot (20 µl) of each sample was injected into the API4000 LC-MS/MS system for analysis, and transitions of 352.05 AMU (Q1) and 267.10 AMU (Q3) were monitored. All pharmacokinetic parameters were derived from concentration–time data by non-compartmental analyses. Pharmacokinetic parameters were calculated using the computer program WinNonlin (Version 6.4) purchased from Certara Company. Results are expressed as mean and standard error of the mean. No further statistical analysis was performed.

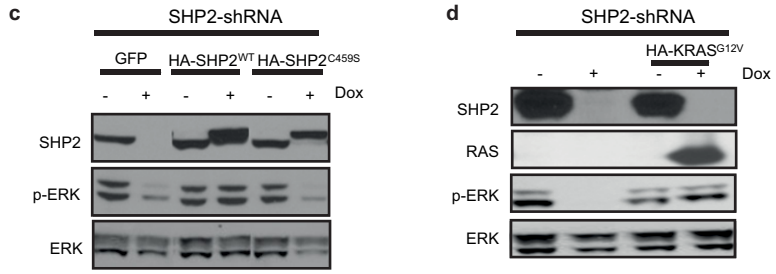
© 2016 Macmillan Publishers Limited. All rights reserved

22. Hoffman, G. R. *et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl Acad. Sci. USA* **111,** 3128–3133 (2014).
23. Shao, D. D. *et al.* ATARiS: computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res.* **23,** 665–678 (2013).
24. Clare, J. J., Tate, S. N., Nobbs, M. & Romanos, M. A. Voltage-gated sodium channels as therapeutic targets. *Drug Discov. Today* **5,** 506–520 (2000).
25. Zhao, H. *et al.* Isoxazole carboxylic acids as protein tyrosine phosphatase 1B (PTP1B) inhibitors. *Bioorg. Med. Chem. Lett.* **14,** 5543–5546 (2004).
26. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010).
27. Bricogne, G. *et al.* BUSTER version 2.8.0. (Global Phasing Ltd., 2009).
28. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
29. Weisberg, E. *et al.* Inhibition of wild-type p53-expressing AML by the novel small molecule HDM2 inhibitor CGM097. *Mol. Cancer Ther.* **14,** 2249–2259 (2015).

© 2016 Macmillan Publishers Limited. All rights reserved

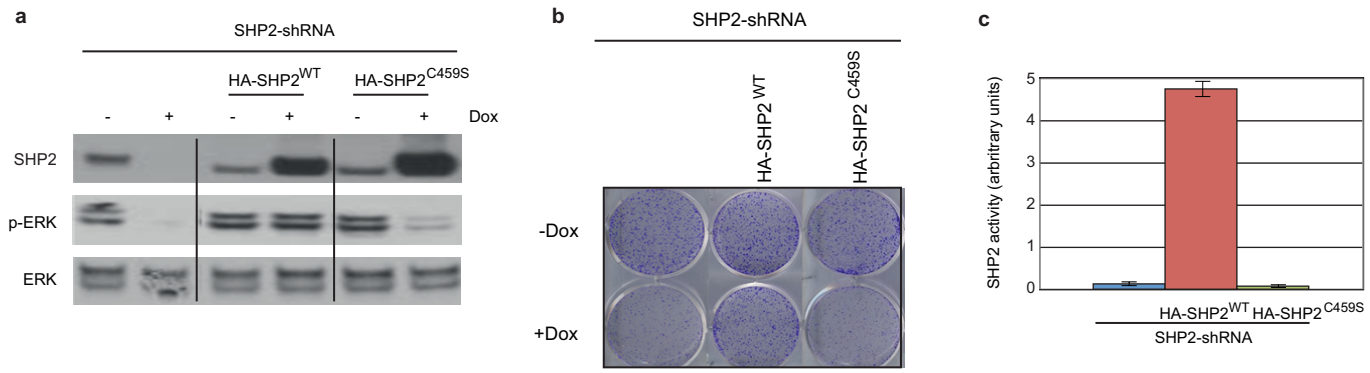






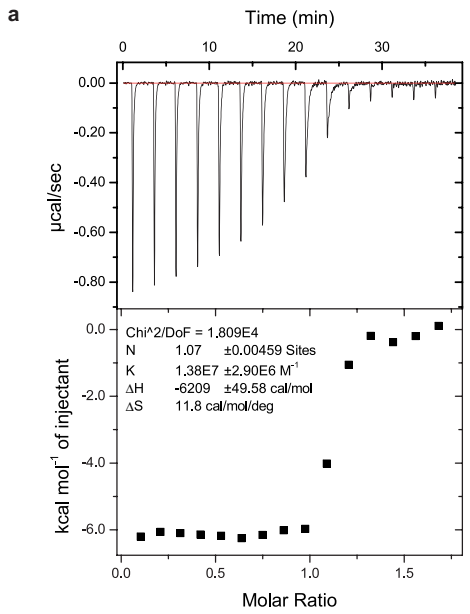
Extended Data Figure 1 | SHP2 depletion inhibits the growth of RTK-amplified cancer cells. a, Cells expressing dox-inducible SHP2 shRNA in various RTK-amplified cancer cells were generated including SUM52 (FGFR2), KATOIII (FGFR2), MDA-MB-468, KYSE520 (EGFR), NCI-H2190 (HER2), NCI-H2228 (EML4-ALK). **b**, Stable clones of MDA-MB-231 ($KRAS^{G13D}$) and A2058 $BRAF^{V600E}$) cancer cells were established as controls. Cells were treated with dox, and colony formation was measured after 11 days by crystal violet staining. **c**, Western blot showing the expression of SHP2, p-ERK and ERK in the presence (+) or absence (−) of dox in MDA-MB-468 SHP2-depleted cells stably expressing either GFP, wild-type HA–SHP2 or HA–$SHP2^{C459S}$. **d**, Western blot of SHP2, p-ERK and ERK in the presence (+) or absence (−) of dox in SHP2-depleted SUM52 cells expressing vector control or HA–$KRAS^{G12V}$. In **c** and **d**, dox treatment induces depletion of endogenous SHP2 protein and simultaneously expression of the exogenous proteins GFP, wild-type HA–SHP2, HA–$SHP2^{C459S}$ or HA–$KRAS^{G12V}$.

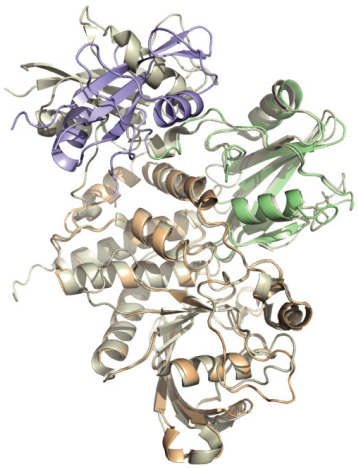
© 2016 Macmillan Publishers Limited. All rights reserved

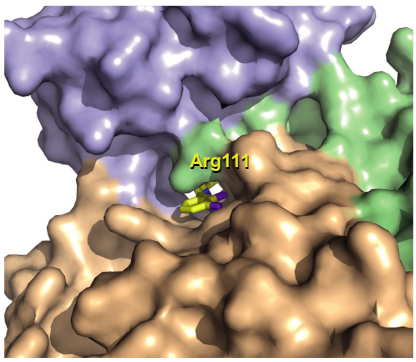


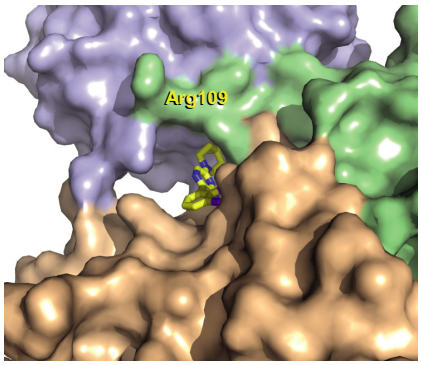
Extended Data Figure 2 | Phosphatase activity is required for cancer growth. a, Western blot of SHP2, p-ERK and ERK in SHP2-depleted SUM52 cells stably expressing vehicle, wild-type SHP2 or HA–SHP2^{C459S}. Note, four lanes corresponding to an unrelated study were removed from the image. All lanes originated from the same gel at the same exposure. **b**, Colony formation of SHP2-depleted SUM52 cells stably expressing vehicle, wild-type SHP2 or HA–SHP2^{C459S}. SHP2 knockdown and SHP2 variant re-expression perform using dox treatment. Colony formation was monitored after 11 days with crystal violet staining. **c**, Phosphatase activity of SHP2-depleted SUM52 cells stably expressing HA–SHP2 or HA–SHP2^{C459S}. Cells were treated with dox for 3 days. SHP2 protein was immunoprecipitated from cell lysates and phosphatase activity was measured using DiFMUP assay. Data are presented as mean ± s.d. ($n = 3$).

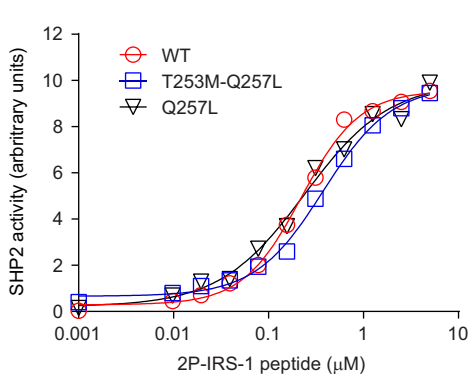
© 2016 Macmillan Publishers Limited. All rights reserved







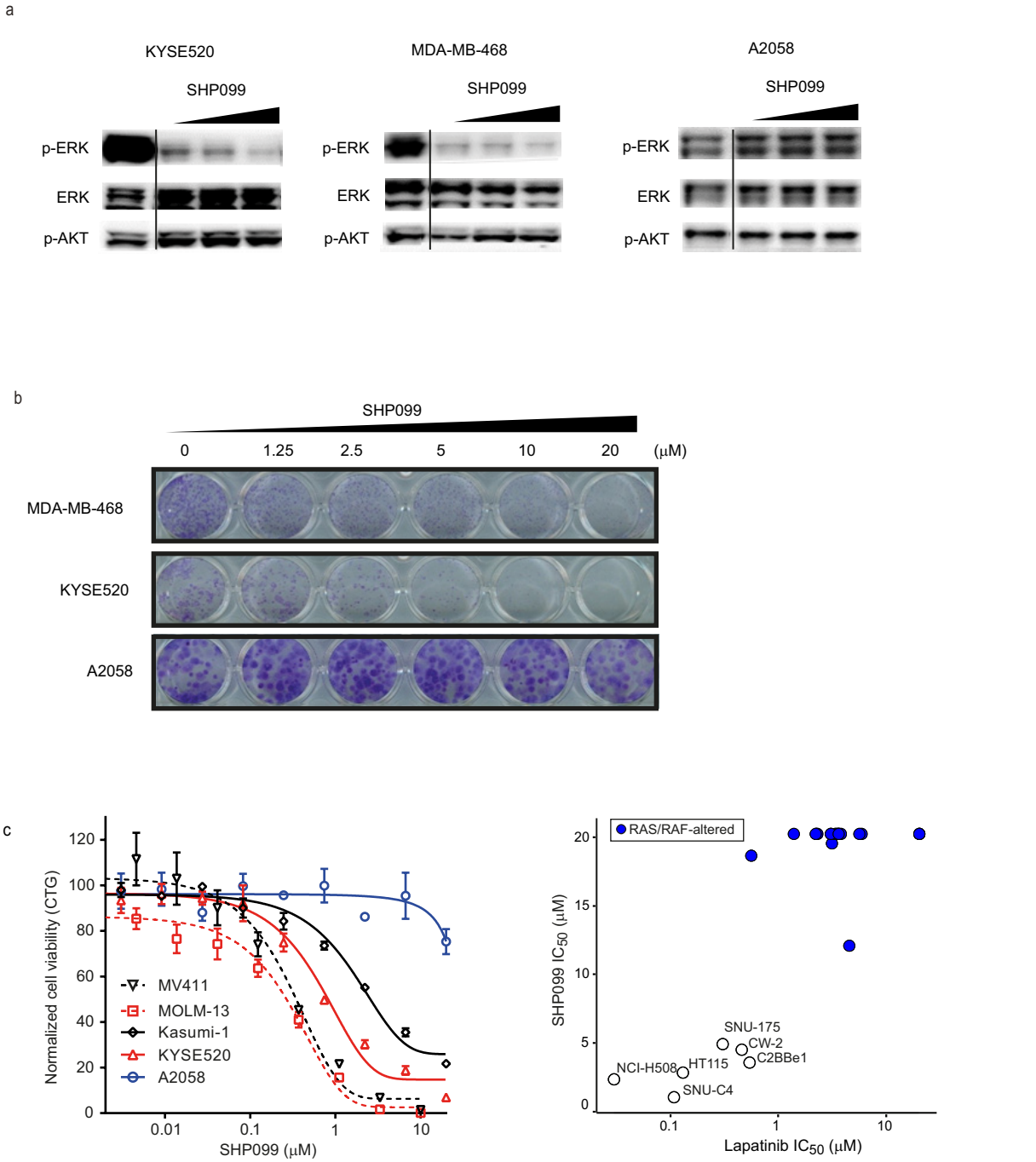




SHP2 variant	2P-IRS-1 activation AC_{50} (µM)	SHP099 IC_{50} (µM)
WT	0.22	0.071
Q257L	0.24	1.33
T253M/Q257L	0.38	> 100

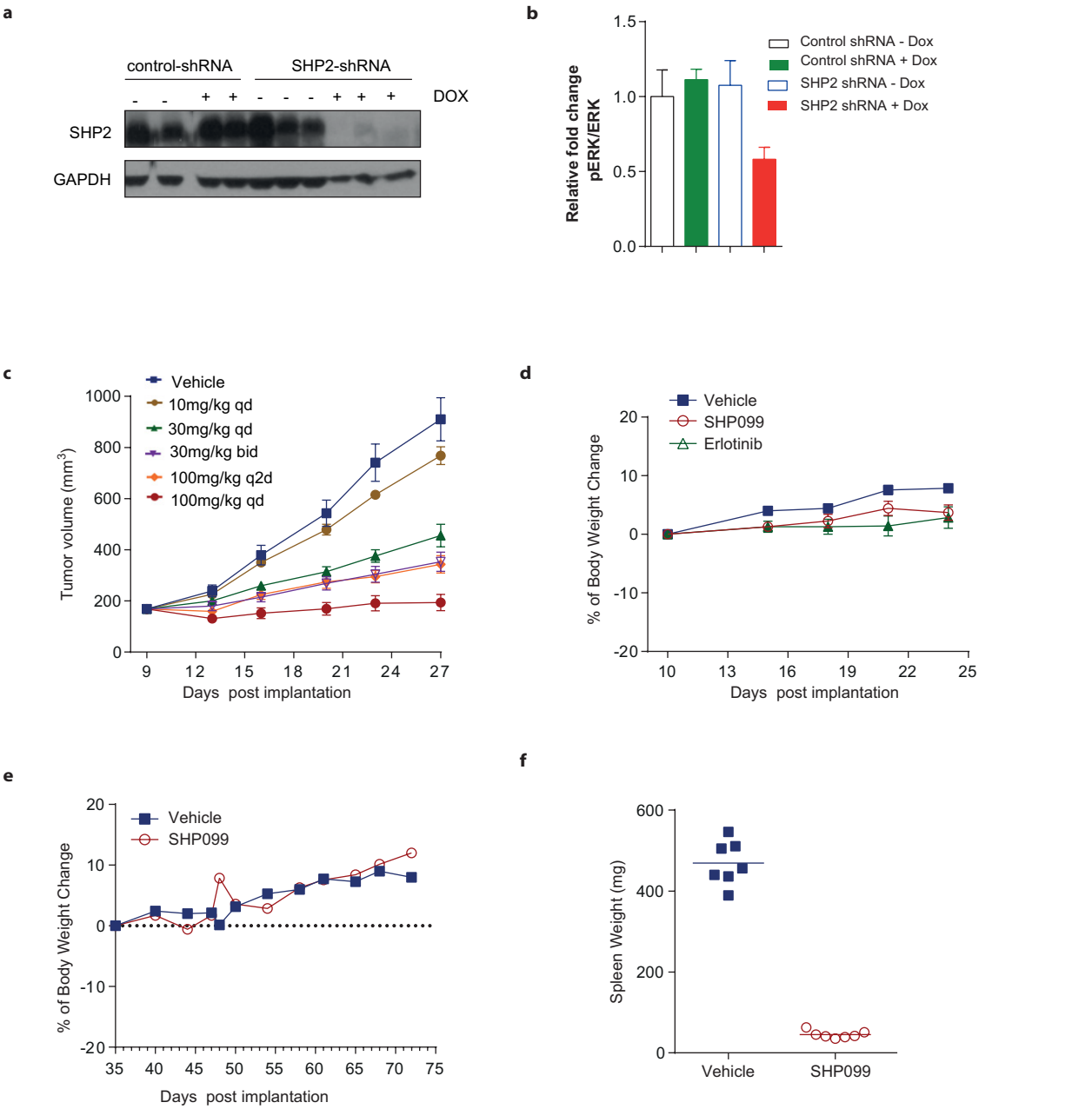
Extended Data Figure 3 | Thermodynamic characterization of the SHP099–SHP2 binding complex, comparison of SHP2 and SHP1's allosteric pocket and characterization of SHP099-resistant SHP2 mutants. a, Isothermal titration calorimetry of SHP099 binding to SHP2. SHP099 binds stoichiometrically to SHP2 with a dissociation constant measured at 73 ± 15 nM. **b**, Structural differences in the central tunnel between SHP1 and SHP2. Ribbon representation of SHP2 (multi-colour) and SHP1 (grey) X-ray structures in the closed conformation. The PTP (tan) and N-SH2 (green) domain overlay well (r.m.s.d. < 1.5 Å), however the C-SH2 (blue) domain has a significantly different orientation. **c**, Surface representation of SHP2–SHP099 co-crystal structure. **d**, Surface representation of SHP1 with SHP099 modelled on the basis of the SHP2 superimposition. Central tunnel is significantly larger in SHP1 owing to a change in orientation of the C-SH2 domain. This change repositions the linker between the two SH2 domains removing several key interactions, highlighted by residue Arg109 in SHP1 and by Arg111 in SHP2 (equivalent residues). **e**, Biochemical activity of wild-type SHP2, $SHP2^{Q257L}$ and $SHP2^{T253M/Q257L}$. SHP2 activity was determined using DiFMUP in the presence of various concentrations of 2P-IRS-1. Data points along the line represent the mean of two replicate values. $SHP2^{Q257L}$ and $SHP2^{T253M/Q257L}$ retain activity regulation and 2P-IRS-1 activation potential comparable to wild-type SHP2 but are 18- and <1,000-fold less sensitive to SHP099 inhibition.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Cellular activity of SHP099. a, Western blot of SHP2, p-ERK, ERK and p-AKT from KYSE520, MDA-MB-468 or A2058 cells treated with SHP099 (1, 3, 10 μM). Note, three lanes corresponding to an unrelated study were removed from the image. All lanes originated from the same gel at the same exposure. **b**, Colony formation of KYSE520, MDA-MB-468, and A2058 in the presence of SHP099. Colony formation was measured after 11 days of SHP099 treatment by crystal violet staining. **c**, SHP099 inhibitory activity against cell lines KYSE520 (EGFR-amplified), MV-411(FLT3-ITD), MOLM-13 (FLT3-ITD), Kasumi (c-Kit altered) and negative control A2058 ($BRAF^{V600E}$) treated with SHP099 concentration varying from 0.0046 to 20 μM. Cellular viability was measured using CellTiter-Glo. Data presented as mean ± s.d. ($n = 3$). **d**, Comparison of SHP099 activity with Lapatinib in a panel of 26 colorectal cell lines. SHP099 sensitivity correlates with sensitivity to Lapatinib, a potent tyrosine kinase inhibitor against Her1/2 and EGFR. RAS- and BRAF-mutated cell lines are shown in blue. Cellular viability was measured using CellTiter-Glo. The corresponding data and cell line genotypes are included in Supplementary Information Table 2.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | SHP2 depletion or inhibition by SHP099 assessed in *in vivo* xenograft models. a, Western blot of SHP2 and GAPDH in KYSE520 xenograft lysates following 14 days of dox treatment (+). **b**, Response of p-ERK in tumour xenograft lysates following 14 days of dox treatment. **c**, Antitumour efficacy of SHP099 administered orally for 14 consecutive days at the doses and schedules indicated. Data are plotted as the treatment mean ± s.e.m. ($n = 9$). **d**, Body weight of mice bearing subcutaneous KYSE520 xenografts and administered either SHP099 (100 mg per kg daily), erlotinib (80 mg per kg daily) or vehicle for 14 consecutive days. Data are presented as treatment mean ± s.e.m. ($n = 7$). **e**, Body weight of mice bearing orthotopic primary tumour derived xenografts and administered an oral gavage of SHP099 at 75 mg per kg daily. Data are presented as treatment mean ± s.e.m. ($n = 7$). **f**, The mouse spleen weight measurement of mice bearing the orthotopic AML xenograft model following 34 days of once-daily dosing with SHP099 at 75 mg per kg. Data are presented as treatment mean ± s.e.m. ($n = 7$).

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Selectivity profiling of SHP099 in phosphatase enzyme panel

Eurofins Phosphatase	Species	Native phosphatase sequence present in recombinant protein	IC_{50}, μM
CD45(h)	Human	598-end	>100
DUSP22(h)	Human	Full length	>100
HePTP(h)	Human	22-end	>100
LMPTP-A(h)	Human	Full length; Q106R	>100
LMPTP-B(h)	Human	Full length	>100
MKP5(h)	Human	320-end	>100
PP1a(h)	Human	Full length	>100
PP2A(h)	Human	Native enzyme	>100
PP5(h)	Human	Full length	>100
PTPb(h)	Human	1643-end	>100
PTP-1B(h)	Human	1-321	>100
PTP MEG1(h)	Human	423-end	>100
PTP MEG2(h)	Human	283-end	>100
PTPN22(h)	Human	1-312	>100
RPTPm(h)	Human	879-1184	>100
SHP-1(h)	Human	Full length	>100
SHP-2(h)	Human	230-545	>100
TCPTP(h)	Human	1-341	>100
TMDP(h)	Human	Full length	>100
VHR(h)	Human	Full length	>100
YopH(Yersinia)	Yersinia	Full length; R211A	>100

Assay was performed using PhosphataseProfiler at Eurofins.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 2 | Selectivity profiling of SHP099 in kinase enzyme panel

Kinases	IC_{50}, μM	Kinases	IC_{50}, μM
CE ABL1 (64-515) nonphos v2	>10	CE MAPK1	>10
CE ACVR1 (172-499)	>10	CE MAPK10	>10
CE AKT1	>10	CE MAPK14	>10
CE ALK (1066-1459)	>10	CE MAPKAPK2	>10
CE AURKA	>10	CE MAPKAPK5 (2-472)	>10
CE BTK	>10	CE MET (956-1390)	>10
CE CAMK2D	>10	CE MKNK1	>10
CE CDK1B	>10	CE MKNK2	>10
CE CDK2A	>10	CE PAK2	>10
CE CDK4D1	>10	CE PDGFRa (551-V561D-1089)	>10
CE CSK	>10	CE PDPK1	>10
CE CSNK1G3 (35-362)	>10	CE PIM2	>10
CE EGFR (668-1210)	>10	CE PKN1	>10
CE EPHB4 (566-987)	>10	CE PKN2	>10
CE ERBB4 (673-1308)	>10	CE PLK1	>10
CE FGFR1 (407-822)	>10	CE PRKACA	>10
CE FGFR2 (406-821)	9.7	CE PRKCA	>10
CE FGFR3 (411-806)	>10	CE PRKCQ	>10
CE FGFR3 (411-K650E-806)	>10	CE RET (658-1072)	>10
CE FGFR4 (388-802)	>10	CE ROCK2 (6-553)	>10
CE FLT3 (563-D835Y-993)	>10	CE RPS6KB1 (1-421)	>10
CE GSK3B	9.9	CE SRC (1-535)	>10
CE INSR (871-1343)	>10	CE STK17B	>10
CE IRAK1 (184-712)	>10	CE SYK (2-635) nonphos	>10
CE IRAK4 (1-460)	>10	CE WNK1 (2-491)	>10
CE JAK1 (866-1154)	>10	CE ZAP70	>10
CE JAK2 (808-1132)	>10	ADP-FRET PIK3CD	>10
CE KDR (807-1356)	>10	ADP-FRET PIK3CG	>10
CE KIT (544-976)	>10	ATP-binding MTOR(1360-2549)	>10
CE LCK (1-508)	>10	KGlo PIK3C3	>10
CE LYN (1-512)	>10	KGlo PIK3CA	>10
CE MAP3K8 (30-404)	>10	KGlo PIK3CB	>10
CE MAP4K4	>10	KGlo PIK4CB	>10

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 3 | Preclinical safety pharmacology off-target activity panel

Assay	AC_{50} (μM)	
	Antagonism	Agonism
5HT2C	>30	N/A
Ad1	>30	N/A
Ad2A	>30	N/A
Ad3	>30	N/A
alpha2B	>30	N/A
alpha2C	>30	N/A
beta1	>30	N/A
AT1	>30	N/A
CCKa	>30	N/A
D2	>30	N/A
D3	>30	N/A
ETa	>30	N/A
GHS	>30	N/A
H1	>30	N/A
H3	>30	N/A
MC3	>30	N/A
h Motilin	>30	N/A
M1	>30	N/A
M3	>30	N/A
op-delta	>30	N/A
op-mu	>30	N/A
hr V1a	>30	N/A
Nic(ns)	>30	N/A
5HT3	6.7	N/A
AdT	>30	N/A
DAT	>30	N/A
NET	>30	N/A
5HTT	>30	N/A
COX-1	>30	N/A
COX-2	14	N/A
MAO-A	>30	N/A
h PDE3	>30	N/A
5HT1A	> 30	> 30
5HT2A	> 30	> 30
5HT2B	> 30	> 30
Alpha 1A	> 30	> 30
Alpha 2A	> 30	> 30
Beta2	> 30	N/A
CB1	> 30	> 30
D1	> 30	N/A
GABAA	> 30	> 30
M2	> 30	> 30
rr AR	> 30	N/A
ERalpha	12	> 10
GR	> 30	> 30
PPARg	> 30	> 30
PR-B	> 30	> 30
PXR	> 30	> 30

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 4 | Crystallographic Data and Refinement Statistics

Parameters	SHP2/SHP099 complex
Space group	$P2_1$
Unit Cell (Å)	a=46.19, b=213.79, c=55.89
Resolution range (Å)	24.89 – 1.70 (1.74 – 1.70)
Total observations	371305
Unique reflections	114992
Completeness (%)[a]	97.8 (87.3)
Multiplicity	3.2 (2.4)
$<I/\sigma(I)>$[a]	17.5 (2.4)
R_{merge}[a,b]	0.034 (0.373)
R_{cryst}/R_{free}[c]	0.195/0.221 (0.235/0.257)
Protein atoms	7768
Heterogen atoms	76
Solvent molecules	670
Average *B*-factor (Å^2)	27.4
R.m.s.d. bond lengths (Å)	0.01
R.m.s.d. bond angle (°)	1.02
Ramachandran Plot (%)	
Favored	97.1
Allowed	2.3
Outliers	0.6

[a]Numbers in parentheses are for the highest resolution shell.
[b]$R_{merge} = \Sigma|I_h - <I_h>|/\Sigma I_h$ over all h, where I_h is the intensity of reflection h.
[c]R_{cryst} and $R_{free} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, where F_o and F_c are observed and calculated amplitudes, respectively. R_{free} was calculated using 5% of data excluded from the refinement.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18629

Inflammasome-activated gasdermin D causes pyroptosis by forming membrane pores

Xing Liu[1,2]*, Zhibin Zhang[1,2]*, Jianbin Ruan[1,3]*, Youdong Pan[4], Venkat Giri Magupalli[1,3], Hao Wu[1,3] & Judy Lieberman[1,2]

Inflammatory caspases (caspases 1, 4, 5 and 11) are activated in response to microbial infection and danger signals. When activated, they cleave mouse and human gasdermin D (GSDMD) after Asp276 and Asp275, respectively, to generate an N-terminal cleavage product (GSDMD-NT) that triggers inflammatory death (pyroptosis) and release of inflammatory cytokines such as interleukin-1β[1,2]. Cleavage removes the C-terminal fragment (GSDMD-CT), which is thought to fold back on GSDMD-NT to inhibit its activation. However, how GSDMD-NT causes cell death is unknown. Here we show that GSDMD-NT oligomerizes in membranes to form pores that are visible by electron microscopy. GSDMD-NT binds to phosphatidylinositol phosphates and phosphatidylserine (restricted to the cell membrane inner leaflet) and cardiolipin (present in the inner and outer leaflets of bacterial membranes). Mutation of four evolutionarily conserved basic residues blocks GSDMD-NT oligomerization, membrane binding, pore formation and pyroptosis. Because of its lipid-binding preferences, GSDMD-NT kills from within the cell, but does not harm neighbouring mammalian cells when it is released during pyroptosis. GSDMD-NT also kills cell-free bacteria *in vitro* and may have a direct bactericidal effect within the cytosol of host cells, but the importance of direct bacterial killing in controlling *in vivo* infection remains to be determined.

We hypothesized that GSDMD-NT might form pores that permeabilize mammalian membranes during pyroptosis. To examine whether GSDMD-NT oligomerizes, we expressed Flag-tagged mouse GSDMD-NT or GSDMD in HEK293T cells and analysed the lysates by SDS–PAGE and Flag-immunoblot (Extended Data Fig. 1a, b). Under non-reducing conditions, GSDMD-NT migrated as both an ~30 kDa monomer and ≥250 kDa multimer. The multimeric band disappeared under reducing conditions, suggesting that GSDMD-NT oligomerization requires disulfide-cross-linking. Flag–GSDMD migrated mostly as a monomer, but a dimeric band was also formed when reactive sulfhydryl groups were not blocked, suggesting that these dimers formed during lysis. When the same cell lysates were analysed by native gel electrophoresis, high molecular weight oligomers were visualized selectively in cells transfected with Flag–GSDMD-NT (Fig. 1a). To confirm the association of multiple GSDMD-NT subunits in the oligomer, we transfected HEK293T cells with Flag- and haemagglutinin (HA)-tagged GSDMD-NT. Immunoprecipitation with either anti-Flag (Fig. 1b) or anti-HA (Extended Data Fig. 1c) antibodies pulled down both tagged species, confirming that GSDMD-NT self-associates and might form homo-oligomers. When the co-immunoprecipitation was repeated in cells transfected with Flag–GSDMD-NT, Flag–GSDMD-CT, and/or GSDMD-CT–MYC, the two species of GSDMD-CT did not co-precipitate, but Flag–GSDMD-NT associated with MYC-tagged GSDMD-CT (Extended Data Fig. 1d).

Ectopic caspase-11 expression triggers pyroptosis in *GSDMD*-expressing cells[1,3]. To determine whether caspase-11 activates GSDMD cleavage and oligomerization, we co-transfected HEK293T cells, which do not express *GSDMD*, with plasmids encoding Flag–GSDMD and wild-type or enzymatically dead (C254A) caspase-11. We analysed cell death by measuring lactate dehydrogenase release and GSDMD oligomerization using SDS–PAGE and immunoblot, probed for Flag and caspase-11 (Fig. 1c). 60% of Flag–GSDMD-expressing cells co-expressing wild-type, but not mutant, caspase-11 were killed (Extended Data Fig. 1e). Only wild-type caspase-11 generated GSDMD-NT and its oligomer. Similar results were obtained when immortalized mouse bone-marrow-derived macrophages (iBMDMs) stably expressing Flag–GSDMD were electroporated with lipopolysaccharide (LPS) to activate caspase-11 (ref. 4; Fig. 1d, Extended Data Fig. 1f). Thus caspase-11 cleaves GSDMD, causing GSDMD-NT oligomerization and pyroptosis.

We hypothesized that GSDMD-NT oligomers form cell membrane pores that kill cells. Pore-forming proteins often use positively charged amphipathic structures for membrane insertion[5–7]. To identify potential functional pore domains, we searched for evolutionarily conserved, positively charged residues in GSDMD-NT, comparing the sequences of six mammalian species using the Clustal Omega and SOPMA secondary structure prediction server[8]. A cluster of four such residues occurs in a pair of predicted amphipathic α-helices (mouse Arg138, Lys146, Arg152, Arg154) (Fig. 1e, upper panel). Because of their possible importance, we engineered mutant forms of mouse Flag–GSDMD-NT containing 2, 3 or 4 Arg or Lys to Ala mutations of these residues. These changes were not predicted to affect the secondary structure (Fig. 1e, lower panel), which was verified by showing that the melting temperatures of wild-type and 4 Ala (4A)-mutated GSDMD-NT were similar (46.8 °C and 45.6 °C, respectively). We also generated a mutant protein in which Arg138 was mutated to Ser. We determined whether these mutations interfere with oligomerization and pyroptosis in HEK293T cells ectopically expressing GSDMD, wild-type or mutated GSDMD-NT (Fig. 1f, g). As expected, wild-type GSDMD-NT, but not GSDMD, triggered both pyroptosis and GSDMD-NT oligomerization. Mutation of all four basic residues completely blocked both pyroptosis and oligomerization, whereas mutations of two or three of the residues resulted in partial blocking. Ectopic Flag- and HA-tagged GSDMD-NT 4A also did not co-immunoprecipitate (Extended Data Fig. 2a). Ala mutations of other conserved basic residues (Lys204, Lys205, Lys237, Arg239), alone or combined with mutations in nonconserved basic residues (Arg248, Lys 249) that were not within predicted amphipathic structures, did not affect pyroptosis (Extended Data Fig. 2b, data not shown). Oligomerization and cell death were correlated, suggesting that GSDMD-NT oligomers were responsible for pyroptosis.

To verify that the 4A mutation inactivates pyroptosis, we knocked down *Gsdmd* in iBMDMs and assessed whether wild-type or 4A-mutant GSDMD restored LPS-transfection-induced pyroptosis (Fig. 1h, Extended Data Fig. 2c, d). *Gsdmd* knockdown strongly inhibited pyroptosis, which was restored by ectopic expression of small interfering RNA (siRNA)-resistant wild-type, but not 4A mutant, *Gsdmd*.

[1]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, Massachusetts 02115, USA. [2]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Dermatology and Harvard Skin Disease Research Center, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved

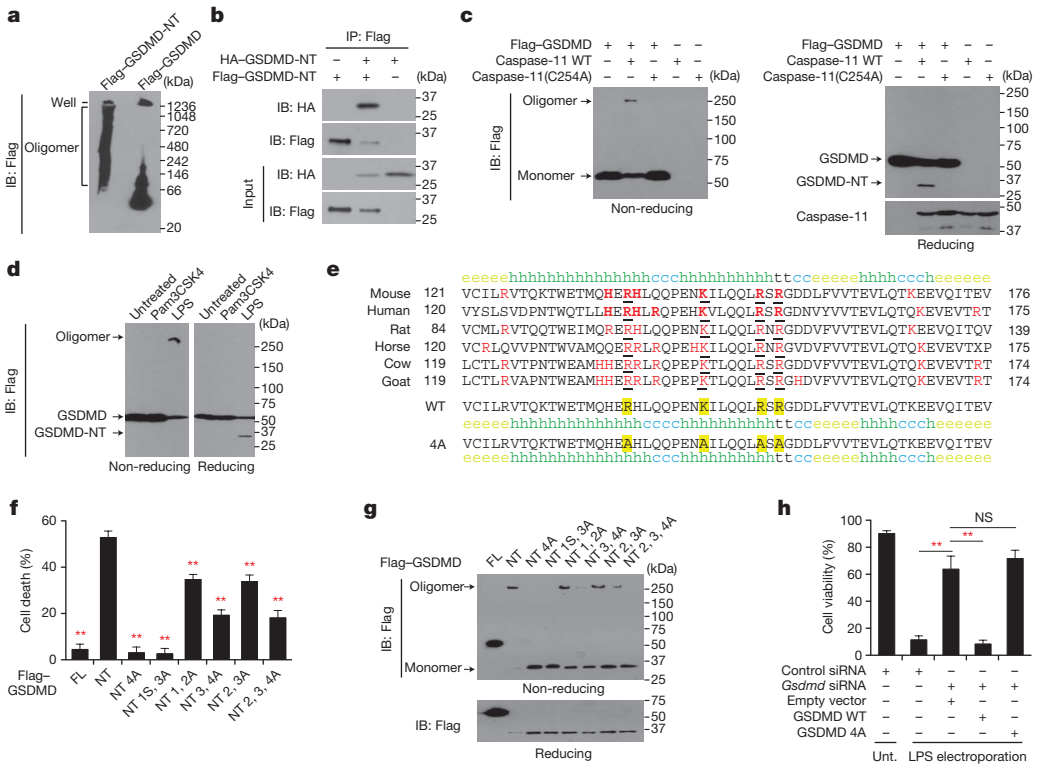


Figure 1 | GSDMD-NT forms oligomers, disrupted by mutation of four positively charged residues. a, HEK293T cells, transfected with indicated plasmids, lysed under non-reducing conditions, were resolved on a native gel, immunoblotted for Flag. **b**, Flag immunoprecipitation of lysates of HEK293T cells, transfected with HA–GSDMD-NT and/or Flag–GSDMD-NT, were analysed by immunoblot. **c**, HEK293T cells, transiently transfected with indicated plasmids, were assessed 16 h after transfection for oligomer formation by Flag immunoblot of non-reducing (left) or reducing (right) gels. The reducing gel was also blotted for caspase-11. **d**, iBMDMs expressing Flag–GSDMD were electroporated with phosphate-buffered solution (PBS), LPS or Pam3CSK4 and analysed 2 h later by Flag immunoblot. **e**, The cluster of four evolutionarily conserved, positively charged amino acids (red and underlined) in GSDMD-NT were mutated to Ala. Secondary structures of the wild-type and 4A-mutated mouse GSDMD fragment were predicted using the SOPMA algorithm. h, helix; e, sheet; t, turn; c, coil. **f**, **g**, Mutations of mouse GSDMD-NT block GSDMD-NT-mediated pyroptosis (**f**) and oligomerization (**g**). The first to fourth amino acids (R138/K146/R152/R154) were mutated to Ala. The mutated residues are indicated, i.e. in NT4A, all 4 residues are mutated; in NT 3, 4A, the 3rd and 4th residues are mutated. In NT1S3A, R138 was mutated to Ser, the other three residues were mutated to Ala. FL, full-length GSDMD; NT, GSDMD-NT. HEK293T cells, transfected with wild-type (WT) or mutated GSDMD-NT, were analysed 18 h later for cell death (CytoTox96 assay) and GSDMD oligomerization (Flag immunoblot). GSDMD-NT monomer and oligomer are indicated. **h**, iBMDMs, co-transfected with control or *Gsdmd* siRNA and the indicated siRNA-resistant *Gsdmd* expression plasmids, were electroporated with LPS. The number of surviving cells was determined by CellTiter-Glo assay 2.5 h later. Mean ± s.d. of three technical replicates from one of three independent experiments are shown (**f**, **h**). Statistical differences are relative to Flag–GSDMD-NT-expressing samples (**f**). $^{**}P < 0.01$ (two-tailed *t*-test). NS, not significant; unt., not transfected with LPS.

Because GSDMD-NT oligomerization was inhibited by reducing agents, we also mutated the six Cys residues in the mouse protein and analysed oligomerization in transfected cells. Mutations of Cys39 or Cys192 impaired oligomerization, suggesting that intramolecular or intermolecular disulfide bonds between these residues are critical for oligomerization (Extended Data Fig. 2e).

If GSDMD-NT forms plasma membrane pores, it should relocalize to the cell membrane after caspase activation. To assess membrane localization, we lysed cells co-transfected with wild-type or 4A Flag–GSDMD and wild-type or C254A caspase-11 in the detergent Triton X-114 to separate cytosolic proteins in the aqueous phase from membrane-associated proteins in the detergent phase[9] (Fig. 2a, b). A cleavage fragment that migrated in the same way as Flag–GSDMD-NT was only produced in cells transfected with wild-type caspase-11. Wild-type and 4A Flag–GSDMD-NT were detected in the aqueous phase, but only wild-type Flag–GSDMD-NT partitioned into the detergent phase and associated with cell membranes.

To determine with which membrane GSDMD-NT associates, we fractionated the post-nuclear supernatant of HEK293T cells, transfected with Flag–GSDMD, or wild-type or 4A GSDMD-NT expression plasmids, into cytosolic (S100), heavy membrane (P7), light membrane (P20) and insoluble cytoplasmic fractions (P100) (Fig. 2c). Flag–GSDMD and 4A GSDMD-NT were solely in the S100 fraction, but Flag–GSDMD-NT fractionated with both the S100 and heavy membrane P7 fraction, which contains plasma-membrane fragments and mitochondria. When HEK293T cells, transfected to express Flag–GSDMD-NT, were separated into soluble and membrane fractions and analysed by immunoblot, the cytosolic fraction contained mostly monomeric Flag–GSDMD-NT, whereas the membrane fraction only contained the high molecular weight oligomer (Fig. 2d). We used confocal immunofluorescence microscopy to visualize the cellular distribution of transiently expressed Flag-tagged GSDMD and wild-type and 4A-mutant GSDMD-NT (Fig. 2e, f). Flag–GSDMD and oligomerization-defective 4A Flag–GSDMD-NT stained the cytosol diffusely, but Flag–GSDMD-NT concentrated on the plasma membrane. Thus, GSDMD-NT oligomerizes in the plasma membrane during pyroptosis.

Lipid binding influences which membranes pore-forming proteins permeabilize. To identify which lipids GSDMD-NT binds, we incubated recombinant GSDMD, GSDMD-NT, GSDMD-CT and 4A GSDMD-NT and the cytotoxic lymphocyte pore-forming proteins, perforin and granulysin, with membranes dotted with different lipids (Fig. 3a, b). Perforin permeabilizes mammalian membranes, whereas granulysin preferentially permeabilizes microbial membranes[10,11]. Consistent with our previous results, GSDMD, GSDMD-CT and 4A GSDMD-NT did not bind to any lipid. GSDMD-NT bound most

© 2016 Macmillan Publishers Limited. All rights reserved

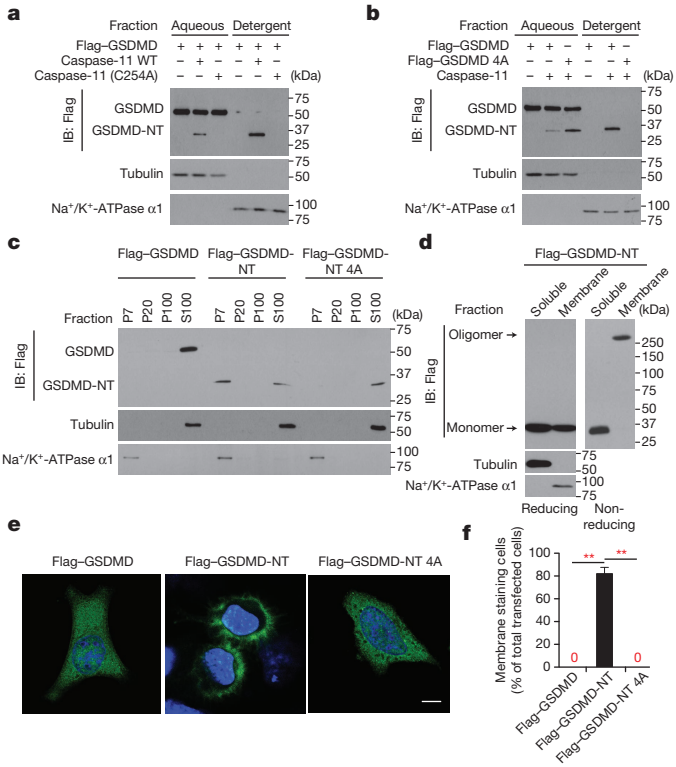


Figure 2 | GSDMD-NT localizes to the plasma membrane. a, **b**, Lysates of HEK293T cells, transfected with indicated plasmids for 16 h, were separated into aqueous and detergent phases using Triton X-114, and analysed by immunoblot probed for Flag, tubulin, or Na^+/K^+-ATPase α1. **c**, HEK293T cells, transfected with indicated plasmids for 16 h, were separated into P7 (heavy membrane), P20 (light membrane), P100 (insoluble cytosol) and S100 (soluble cytosol) fractions and analysed by immunoblot with indicated antibodies. **d**, Soluble and crude membrane fractions of HEK293T cells, transfected to express Flag–GSDMD-NT, were analysed by immunoblot as indicated. **e**, **f**, Representative confocal microscopy images (**e**) and quantification (**f**) of distribution of ectopic Flag–GSDMD, Flag–GSDMD-NT and Flag–GSDMD-NT 4A (green) in HeLa cells co-stained with DAPI (blue). The ratio of cells with membrane versus cytosolic Flag staining was calculated by counting 10 high-power fields for each sample in 5 independent experiments (**f**). $^{**}P < 0.0001$ (paired *t*-test). Scale bar, 20 μm. Data are representative of three independent experiments (**a**–**d**).

strongly to the mitochondrial and bacterial lipid, cardiolipin, and to the phosphatidylinositol phosphates (PIPs), PtdIns(4)P and PtdIns(4,5)P2, and less strongly to phosphatidic acid (PA) and phosphatidylserine (PS), which are all on the mammalian cell membrane inner leaflet[12,13]. It did not bind to phosphatidylethanolamine (PE) or phosphatidylcholine (PC), the major lipids on both plasma membrane leaflets. Cardiolipin in the mitochondrial inner membrane is inaccessible to the cytosol[14]. This binding pattern suggests that GSDMS-NT may selectively bind to the plasma membrane from within and to bacterial membranes. The outer leaflet of endosome and phagosome membranes contains the same phospholipids as the plasma membrane inner leaflet, suggesting that GSDMD-NT may also bind to these organelles. In comparison, perforin bound to PE, but not PS, and the same PIPs as GSDMD-NT, consistent with its role in permeabilizing mammalian cell membranes from the outside; granulysin also bound to cardiolipin, consistent with its role in microbial immunity. Mixed lineage kinase domain-like protein, the pore-forming protein activated during necroptosis, which binds to the inner leaflet of the cell membrane, has a similar binding pattern as GSDMD-NT[15–18].

To confirm lipid binding by GSDMD-NT, we measured wild-type and 4A-mutant GSDMD-NT binding and disruption of PE–PC liposomes containing no added lipid or PtdIns(4)P, PtdIns(4,5)P2, PS or PA (Fig. 3c, d). 4A GSDMD-NT did not bind any of these liposomes, whereas wild-type GSDMD-NT bound to all of the liposomes containing the added phospholipids, but not to the PE–PC liposomes. To measure liposome leakage, PE–PC–PS liposomes were prepared that encapsulated Tb^{3+} ions. Alone, Tb^{3+} is weakly fluorescent, but fluoresces strongly when bound to dipicolinic acid (DPA)[19]. Fluorescence of PE–PC–PS liposomes in DPA-containing solutions sharply increased after adding wild-type, but not mutant, GSDMD-NT, indicating Tb^{3+} leakage. Similarly, PS-containing liposomes became leaky after incubation with caspase-11-treated GSDMD, but not after incubation with caspase-11 or GSDMD alone. Thus GSDMD-NT binds to liposomes containing PS or PIPs and disrupts them. The buffer used for these experiments is Ca^{++}-free, suggesting that GSDMD-NT oligomerization, unlike perforin oligomerization, is Ca^{++}-independent.

We next used negative staining electron microscopy to visualize GSDMD-NT oligomers on PS-containing liposomes. Liposomes incubated with GSDMD and caspase-11 showed ruptured morphology, whereas those incubated with only GSDMD did not (Fig. 3e). The ruptured liposomes were decorated with neck-like structures with ~30 nm diameters at membrane openings, which may represent side views of GSDMD-NT pores. To visualize these potential pore-like structures top-down, we used detergent to extract the reconstituted pores from liposomes and purified the proteins through a size-exclusion column before performing negative staining electron microscopy. Stable ring structures with ~15 nm inner and ~32 nm outer diameters were observed, but only when both caspase-11 and GSDMD were added (Fig. 3f). Cleaved interleukin-1β, released from cells undergoing pyroptosis[20], has a diameter of ~4.5 nm (ref. 21) and could readily pass through these pores.

Pyroptotic cells release cytosolic contents into the surrounding media[22]. We used Flag immunoblot to determine whether pyroptotic HEK293T cells ectopically expressing Flag–GSDMD-NT release GSDMD-NT into the culture medium (Fig. 4a). Whereas ectopic Flag–GSDMD was only detected in the cell, Flag–GSDMD-NT was mostly detected in culture supernatants. To examine the activity of released GSDMD-NT, we assessed iBMDM viability after incubation with five-fold-concentrated culture supernatants from HEK293T cells ectopically expressing Flag–GSDMD-NT or Flag–GSDMD (Extended Data Fig. 3a). Neither supernatant killed iBMDMs. These results were confirmed by examining propidium iodide uptake of CFSE-labelled untransfected HEK293T cells after incubation with Flag–GSDMD-NT-expressing HEK293T cells (Extended Data Fig. 3b). Virtually all of the transfected cells died, but none of the bystander cells, consistent with previous reports[2,23]. Thus GSDMD-NT does not injure bystander cells—it does not disrupt the plasma membrane from the outside, which is expected as it only binds to phospholipids present on the inner leaflet of the plasma membrane of viable cells (Fig. 3a, b).

As GSDMD-NT also binds to cardiolipin, we investigated whether the concentrated pyroptotic cell supernatant kills bacteria (Fig. 4b). The pyroptotic supernatant reduced *Escherichia coli* colonies in a dose-dependent manner. As pyroptotic cell supernatants contain many antibacterial factors, including lysosomal enzymes and lysozyme, we assessed the anti-bacterial activity of culture supernatants that were immunodepleted of Flag–GSDMD-NT (Fig. 4c). Depletion of GSDMD-NT inhibited bacterial killing, supporting a direct antibacterial effect of GSDMD-NT. These culture supernatants were concentrated from cells overexpressing GSDMD-NT, an unphysiological condition that might have unnaturally exaggerated bacterial killing. To examine whether enough endogenous GSDMD-NT is released to kill extracellular bacteria, we collected antibiotic-free culture supernatants from iBMDMs, transfected with LPS or control Pam3CSK4 or treated with LPS and nigericin for 3 h, and added them at dilutions of 1:4 or 1:2 to *Listeria monocytogenes*. Addition of unconcentrated pyroptotic iBMDM culture supernatants significantly reduced bacterial colony-forming units (c.f.u.) in a dose-dependent manner (Fig. 4d).

© 2016 Macmillan Publishers Limited. All rights reserved

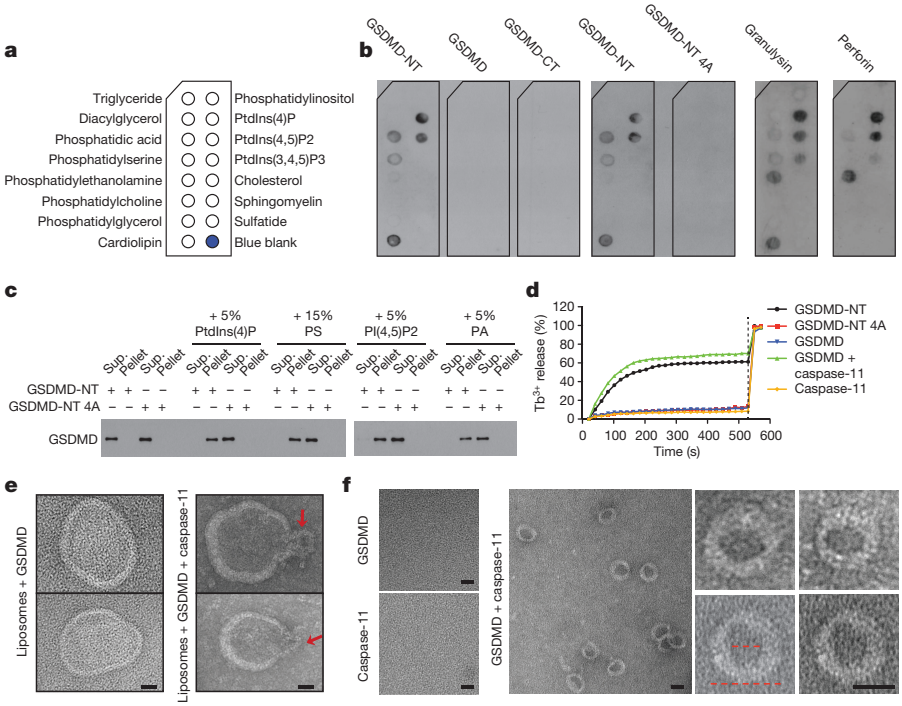


Figure 3 | N-terminal gasdermin D binds to phosphatidyl serine and cardiolipin and forms pores on liposomes. a, **b**, Membranes displaying lipids (**a**) were incubated with indicated proteins and binding was assessed by blotting for GSDMD, perforin or granulysin (**b**). **c**, Wild-type or 4A-mutant GSDMD-NT binding to PC–PE liposomes containing additional indicated phospholipids (molar proportion of added lipid indicated) was analysed by SDS–PAGE and GSDMD immunoblot. **d**, Liposome leakage was monitored by terbium (Tb^{3+}) fluorescence after incubation with the indicated GSDMD protein ± caspase-11. Detergent was added after 9 min (dotted line). **e**, Negative staining electron microscopy images of PE–PC–PS liposomes treated with GSDMD (left) or caspase-11-activated GSDMD (right). Arrows indicate potential side views of GSDMD-NT pores. **f**, Negative staining electron microscopy images of GSDMD-NT pores formed in PS-containing liposomes and extracted by detergent. The left image of pores formed by GSDMD and caspase-11 shows a field with multiple rings, whereas the right images show enlarged single rings. The inner and outer diameters (red dotted lines) are approximately 15 nm and 32 nm, respectively. Data are representative of three independent experiments. Scale bars, 20 nm

To confirm that GSDMD-NT accounts for the anti-bacterial activity, we measured *E. coli* and *Staphylococcus aureus* c.f.u. after incubation with nanomolar concentrations of recombinant GSDMD, GSDMD-CT, wild-type or 4A GSDMD-NT, or granulysin (Fig. 4e). Wild-type GSDMD-NT strongly inhibited colony formation of both bacteria, but the other GSDMD constructs had no anti-bacterial activity. Moreover GSDMD-NT was more active than granulysin. The anti-bacterial effect was rapid: after only 5 min, bacterial c.f.u. were reduced ~2-fold (Fig. 4f). Bacterial growth measurements after treatment with the GSDMD proteins confirmed these results (Extended Data Fig. 3c). To determine whether GSDMD-NT is bactericidal, we measured propidium iodide uptake by *E. coli* and *L. monocytogenes* after treatment for 20 min with the same GSDMD constructs (Extended Data Fig. 3d, data not shown). Wild-type GSDMD-NT killed ~80% of bacteria, but 4A GSDMD-NT, GSDMD-CT and GSDMD had no effect. We next used spinning disk fluorescence microscopy to visualize whether AlexaFluor-488-labelled GSDMD-CT or GSDMD, treated or not with caspase-11, bound to mCherry-expressing *L. monocytogenes* (Extended Data Fig. 3e). Only caspase-11-treated GSDMD bound. Thus GSDMD-NT, released from pyroptotic cells, rapidly binds to and kills both Gram-negative and Gram-positive bacteria.

Intracellular bacteria trigger pyroptosis when LPS on cytosolic Gram-negative bacteria activates the noncanonical inflammasome or when invasive Gram-negative or -positive bacteria activate the canonical inflammasome[4,24–28]. We first looked at whether ectopic GSDMD or wild-type or 4A-mutant GSDMD-NT kills intracellular *L. monocytogenes* in HeLa cells (Fig. 4g). HeLa cells were infected 6 h after transfection. Although expression of GSDMD or 4A GSDMD-NT had no effect, wild-type GSDMD-NT significantly reduced recovery of viable bacteria 6 h and 12 h later. To assess whether cleavage of endogenous GSDMD induces intracellular bacterial killing, we examined the effect of inflammasome activation on the survival of intracellular *L. monocytogenes* in iBMDMs. When LPS-primed-iBMDMs infected with *L. monocytogenes* were treated with nigericin for 1 h to activate the canonical inflammasome, bacterial c.f.u. were reduced ~2-fold (Fig. 4h). Infection of iBMDMs with *L. monocytogenes* also independently triggers AIM2/ASC/caspase-1-mediated pyroptosis[28]. To assess the importance of GSDMD-NT bacterial killing by direct listerial inflammasome activation, we examined the effect of *Gsdmd* knockdown on bacterial c.f.u. (Fig. 4i). iBMDMs with *Gsdmd* knocked down contained three-fold more bacteria, indicating that inflammasome activation in infected cells causes GSDMD-dependent death of intracellular bacteria. The intracellular infection experiments (Fig. 4g–i) were performed without antibiotics, but similar results were obtained when gentamicin was used to kill extracellular bacteria and removed before triggering pyroptosis (data not shown). Thus, inflammasome activation of GSDMD kills both intracellular and extracellular bacteria *in vitro*. However, viable bacteria were not completely eliminated from these cultures. GSDMD-NT could reduce intracellular bacteria by causing host cell death, expelling bacteria from the intracellular niche that is favourable for their survival and replication[29], or by a direct anti-bacterial effect. We have no experimental method to dissociate eukaryotic cell death from bacterial cell death at this time.

How inflammatory caspase cleavage of GSDMD causes pyroptosis[1,2] was previously unknown. Here we show that GSDMD-NT binds to membranes containing PS, cardiolipin, or PIPs to form oligomeric pores that kill mammalian cells and the bacteria that trigger pyroptosis. GSDMD-NT is released into the extracellular milieu during pyroptosis. Because GSDMD-NT binds selectively to phospholipids that are restricted to the inner leaflet of mammalian cell membranes,

© 2016 Macmillan Publishers Limited. All rights reserved

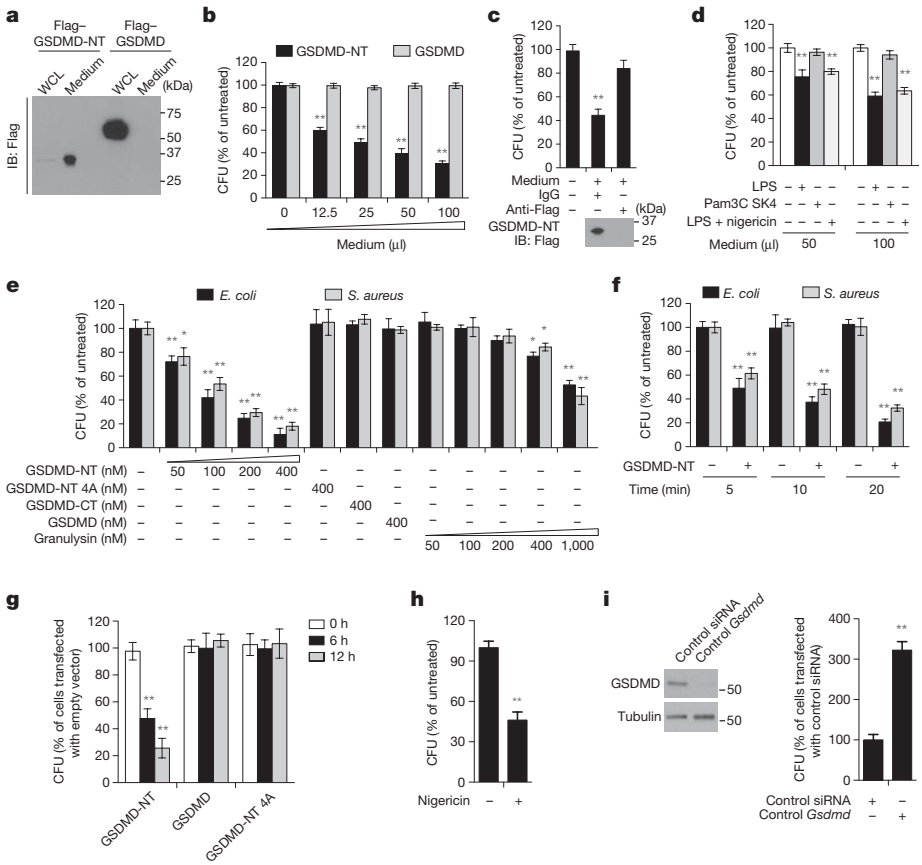


Figure 4 | GSDMD-NT kills bacteria. a, Culture supernatants and whole-cell lysates (WCL) of HEK293T cells, transiently expressing Flag–GSDMD-NT or Flag–GSDMD for 20 h, were analysed by Flag immunoblot of reducing gel. **b**, Antibiotic-free culture supernatants (concentrated fivefold) from transfected HEK293T cells, collected 30 h after transfection, were added to *E. coli*, which were cultured at 37 °C in 200 μl final volume for 30 min before measuring c.f.u. **c**, The concentrated culture medium from Flag–GSDMD-NT-expressing HEK293T cells was immunodepleted with anti-Flag or control IgG, before adding to *E. coli*, as in **b**. Lower panel, Flag immunoblot. **d**, *L. monocytogenes* were incubated at 37 °C for 30 min with antibiotic-free culture supernatants from iBMDMs, transfected with LPS or Pam3CSK4 or incubated with LPS and nigericin for 3 h, before assessing c.f.u. **e**, **f**, *E. coli* and *S. aureus* were treated with the indicated proteins for 20 min (**e**) or with 200 nM wild-type GSDMD-NT for the indicated times (**f**) before measuring c.f.u. **g**, HeLa cells, transfected for 6 h to express the indicated proteins, were infected with *L. monocytogenes* for the indicated times before cells were lysed to analyse intracellular c.f.u. **h**, LPS-primed-iBMDMs infected with *L. monocytogenes*, were treated or not with nigericin for 1 h before bacteria were collected and c.f.u. was analysed. Nigericin had no effect on cell-free bacteria (not shown). **i**, iBMDMs, transfected with control or *Gsdmd* siRNA, were infected with *L. monocytogenes* and assessed for intracellular c.f.u. 12 h later. GSDMD immunoblot, left. Shown are mean ± s.d. of triplicates of one experiment of three (**b**, **d**–**f**, **h**) or two (**c**, **g**, **i**) independent experiments. Statistical differences compared to untreated control samples (two-tailed *t*-test); $*P < 0.05$, $**P < 0.01$.

GSDMD-NT does not kill bystander cells. This selective activity should control tissue damage. GSDMD-NT killing of intracellular bacteria should limit the release of viable bacteria from pyroptotic cells and reduce the spread of infection. We do not know whether GSDMD-NT is active only against bacteria that have escaped from the phagosome. Because the phagosome outer leaflet derives from the inner leaflet of the plasma membrane, phagosomes could be targeted by GSDMD-NT, providing a mechanism for lysis of bacteria within phagosomes. Released GSDMD-NT is active on extracellular bacteria, which probably also helps to control infection. *In vivo* experiments to show that GSDMD-mediated bacterial pore formation protects against bacterial infection would be needed to determine whether direct bacterial killing is physiologically important. However, we do not have a way to distinguish *in vivo* direct killing of bacteria from killing of the host cell (pyroptosis), as the mechanisms that disrupt one also disrupt the other.

A better understanding of how GSDMD-NT forms pores and a more complete description of the GSDMD-NT pore could be obtained by solving the structures of monomeric and oligomerized GSDMD-NT. The oligomers formed in cells overexpressing GSDMD-NT on native gels (Fig. 1a) appeared to be heterogeneous in size, whereas the purified reconstituted pores (Fig. 3f) appeared homogeneous. Direct visualization of the pores formed on cellular membranes should determine whether the pores are uniform. Our identification of mutations that inactivate pore formation, but probably do not affect its overall structure, should help to assess the importance of GSDMD-NT pores in controlling *in vivo* infection.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 March; accepted 9 June 2016.

1. Shi, J. *et al.* Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature* **526,** 660–665 (2015).
2. Kayagaki, N. *et al.* Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature* **526,** 666–671 (2015).
3. Rathinam, V. A. *et al.* TRIF licenses caspase-11-dependent NLRP3 inflammasome activation by gram-negative bacteria. *Cell* **150,** 606–619 (2012).
4. Hagar, J. A., Powell, D. A., Aachoui, Y., Ernst, R. K. & Miao, E. A. Cytoplasmic LPS activates caspase-11: implications in TLR4-independent endotoxic shock. *Science* **341,** 1250–1253 (2013).
5. Law, R. H. *et al.* The structural basis for membrane binding and pore formation by lymphocyte perforin. *Nature* **468,** 447–451 (2010).
6. Montal, M. Design of molecular function: channels of communication. *Annu. Rev. Biophys. Biomol. Struct.* **24,** 31–57 (1995).
7. Rosado, C. J. *et al.* The MACPF/CDC family of pore-forming toxins. *Cell. Microbiol.* **10,** 1765–1774 (2008).

© 2016 Macmillan Publishers Limited. All rights reserved

8. Geourjon, C. & Deléage, G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **11,** 681–684 (1995).
9. Bordier, C. Phase separation of integral membrane proteins in Triton X-114 solution. *J. Biol. Chem.* **256,** 1604–1607 (1981).
10. Dotiwala, F. *et al.* Killer lymphocytes use granulysin, perforin and granzymes to kill intracellular parasites. *Nat. Med.* **22,** 210–216 (2016).
11. Walch, M. *et al.* Cytotoxic cells kill intracellular bacteria through granulysin-mediated delivery of granzymes. *Cell* **157,** 1309–1323 (2014).
12. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9,** 112–124 (2008).
13. Leventis, P. A. & Grinstein, S. The distribution and function of phosphatidylserine in cellular membranes. *Annu. Rev. Biophys.* **39,** 407–427 (2010).
14. Schlame, M. Cardiolipin synthesis for the assembly of bacterial and mitochondrial membranes. *J. Lipid Res.* **49,** 1607–1620 (2008).
15. Pasparakis, M. & Vandenabeele, P. Necroptosis and its role in inflammation. *Nature* **517,** 311–320 (2015).
16. Dondelinger, Y. *et al.* MLKL compromises plasma membrane integrity by binding to phosphatidylinositol phosphates. *Cell Reports* **7,** 971–981 (2014).
17. Wang, H. *et al.* Mixed lineage kinase domain-like protein MLKL causes necrotic membrane disruption upon phosphorylation by RIP3. *Mol. Cell* **54,** 133–146 (2014).
18. Sun, L. *et al.* Mixed lineage kinase domain-like protein mediates necrosis signaling downstream of RIP3 kinase. *Cell* **148,** 213–227 (2012).
19. Wilschut, J. & Papahadjopoulos, D. Ca^{2+}-induced fusion of phospholipid vesicles monitored by mixing of aqueous contents. *Nature* **281,** 690–692 (1979).
20. He, W. T. *et al.* Gasdermin D is an executor of pyroptosis and required for interleukin-1β secretion. *Cell Res.* **25,** 1285–1298 (2015).
21. Finzel, B. C. *et al.* Crystal structure of recombinant human interleukin-1β at 2.0 Å resolution. *J. Mol. Biol.* **209,** 779–791 (1989).
22. Lamkanfi, M. & Dixit, V. M. Mechanisms and functions of inflammasomes. *Cell* **157,** 1013–1022 (2014).
23. Rühl, S. & Broz, P. Caspase-11 activates a canonical NLRP3 inflammasome by promoting K^+ efflux. *Eur. J. Immunol.* **45,** 2927–2936 (2015).
24. Warren, S. E., Mao, D. P., Rodriguez, A. E., Miao, E. A. & Aderem, A. Multiple Nod-like receptors activate caspase 1 during *Listeria monocytogenes* infection. *J. Immunol.* **180,** 7558–7564 (2008).
25. Cervantes, J., Nagata, T., Uchijima, M., Shibata, K. & Koide, Y. Intracytosolic *Listeria monocytogenes* induces cell death through caspase-1 activation in murine macrophages. *Cell. Microbiol.* **10,** 41–52 (2008).
26. Aachoui, Y. *et al.* Caspase-11 protects against bacteria that escape the vacuole. *Science* **339,** 975–978 (2013).
27. Wu, J., Fernandes-Alnemri, T. & Alnemri, E. S. Involvement of the AIM2, NLRC4, and NLRP3 inflammasomes in caspase-1 activation by Listeria monocytogenes. *J. Clin. Immunol.* **30,** 693–702 (2010).
28. Sauer, J. D. *et al.* Listeria monocytogenes triggers AIM2-mediated pyroptosis upon infrequent bacteriolysis in the macrophage cytosol. *Cell Host Microbe* **7,** 412–419 (2010).
29. Miao, E. A. *et al.* Caspase-1-induced pyroptosis is an innate immune effector mechanism against intracellular bacteria. *Nat. Immunol.* **11,** 1136–1142 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by US NIH grant R01AI123265 (J.L.).

Author Contributions X.L. conceived the study. X.L., Z.Z., J.R., H.W. and J.L. designed the experiments and analysed the data. Experiments were performed as follows (X.L., Figs 1a–d, h, 2, 3a–d, 4a, c–f, h, Extended Data Fig. 1a–c, e, f; Z.Z. Figs 1e–g, 3b, d, Fig, 4b, e–h, Extended Data Fig. 1d; J.R. Figs 3, 4e, f; Y.P. Fig. 2e, f; V.M. Fig. 3e). X.L., H.W. and J.L. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.W. (wu@crystal.harvard.edu) or J.L. (judy.lieberman@childrens.harvard.edu).

Reviewer Information *Nature* thanks F. Sigworth and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to outcome assessment.

Cell lines and bacterial strains. HEK293T and HeLa cells were obtained from ATCC, and C57BL/6 mouse iBMDM cells were provided by J. Kagan (Boston Children's Hospital). Cells were cultured in DMEM (Invitrogen) with 10% heat-inactivated foetal bovine serum, supplemented with 100 U ml^{-1} penicillin G, 100 μg ml^{-1} streptomycin sulphate, 6 mM HEPES, 1.6 mM L-glutamine, and 50 μM 2-mercaptoethanol (2ME). There were no antibiotics in the cell culture medium used for bacterial infection and for experiments in which culture supernatants were collected for bacterial incubation. Cells were verified to be free of mycoplasma contamination. Transient transfection of HEK293T and HeLa cells was performed using the calcium phosphate method or Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. iBMDM cells were transfected by nucleofection (Amaxa) using the Amaxa Nucleofector kit (VPA-1009). Bacterial strains were obtained from ATCC (*E. coli* strain BL21, *S. aureus* strain CA-MRSA USA300 and *L. monocytogenes* 10403S strain) and grown in Luria broth (LB), tryptic soy broth and brain–heart infusion media, respectively.

Reagents. Polyclonal anti-human GSDMD was from Novus Biologicals (NBP2-33422) or Proteintech (20770-1-AP). Monoclonal anti-haemagglutinin (F-7) antibody (sc-7392) was from Santa Cruz Biotechnology. Monoclonal anti-Flag M2 antibody (F1804), monoclonal anti-human α-tubulin antibody (T5168) and monoclonal anti-mouse caspase-11 antibody (C1354) were from Sigma-Aldrich. Polyclonal anti-human Na^+/K^+-ATPase α1 (ATP1A1) antibody (#3010) was from Cell Signaling. c-MYC (9E10) monoclonal antibody (MMS-15P) was from Covance. Monoclonal anti-human perforin antibody (3465-6-250) and polyclonal anti-human granulysin antibody (AF3138) were from Mabtech and Novus, respectively. *N*-ethylmaleimide, 2ME, DTT, terbium(III) chloride, DPA (dipicolinic acid) and nigericin were from Sigma-Aldrich. Ultrapure LPS and Pam3CSK4 were from InvivoGen. The complete protease inhibitor cocktail was from Roche. siRNA duplexes targeting *Gsdmd* (s87492; 5′-GGUGAACAUCGGAAAGAUUTT-3′) and the nonspecific control siRNA (CTL, 4390843) were from Ambion.

Plasmids. pCMV6-*Gsdmd* and pCMV-Flag-caspase-11 constructs were obtained from Origene and Addgene, respectively. cDNA for *Gsdmd* was subcloned into pFlag-CMV4, pcDNA3-N-HA and pcDNA3-C-5xMyc. GSDMD truncation mutants were derived by PCR from the corresponding plasmids. All point mutations were generated using QuikChange XL site-directed mutagenesis (Stratagene). All plasmids were verified by sequencing.

Protein expression and purification. Full-length human GSDMD was cloned into the pDB.His.MBP vector with a tomato etch virus (TEV)-cleavable N-terminal His_6-MBP tag using NdeI and Xho I restriction sites. 4A GSDMD, GSDMD-CT, and wild-type and 4A-mutant GSDMD were constructed by QuikChange Mutagenesis (Agilent Technologies). For expression and purification of full-length GSDMD, GSDMD-CT and GSDMD-NT 4A mutant, *E. coli* BL21 (DE3) cells harbouring the indicated plasmids were grown in LB medium supplemented with 50 μg ml^{-1} kanamycin. Protein expression was induced at 18 °C overnight by 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) when OD_{600} reached 0.8. Cells were collected and resuspended in lysis buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 20 mM imidazole and 5 mM 2ME, and lysates were homogenized by ultrasonication. The cell lysate was clarified by centrifugation at 40,000*g* at 4 °C for 1 h. The supernatant containing the target protein was incubated with Ni-NTA resin (Qiagen) that was pre-equilibrated with lysis buffer for 30 min at 4 °C. After incubation, the resin–supernatant mixture was poured into a column and the resin was washed with lysis buffer. The protein was eluted using the lysis buffer supplemented with 500 mM imidazole. The His_6-MBP-tagged protein was further purified by HiTrap Q ion-exchange and Superdex 200 gel-filtration chromatography (GE Healthcare Life Sciences). The His_6-MBP tag was removed by overnight TEV protease digestion at 16 °C. The cleaved protein was purified using HiTrap Q ion-exchange and Superdex 200 gel-filtration columns.

The yield of wild-type GSDMD-NT was lower than for the other constructs because it inserted into bacterial membranes and killed >50% of bacteria after overnight expression, and thus required a different purification strategy. To purify wild-type GSDMD-NT, cells containing the pDB-His_6-MBP-GSDMD-NT clones were grown and induced as described for full-length GSDMD. They were collected and resuspended in lysis buffer containing 25 mM Tris-HCl (pH 8.0) and 150 mM NaCl, and homogenized by ultrasonication. The membrane fraction was harvested by ultracentrifugation at 200,000*g* at 4 °C for 1 h and resuspended and solubilized with 1.0% *n*-dodecyl-β-D-maltoside (DDM) in lysis buffer supplemented with 20 mM imidazole using a glass homogenizer, followed by centrifugation at 200,000*g* at 4 °C for 45 min. The supernatant containing the solubilized protein was incubated for 30 min at 4 °C with Ni-NTA resin (Qiagen) that was pre-equilibrated with the lysis buffer containing 1% DDM and 20 mM imidazole. The resin was washed with lysis buffer containing 1% DDM and 20 mM imidazole, and the recombinant proteins were eluted with lysis buffer supplemented with 500 mM imidazole and 0.1% DDM. The His_6-MBP-GSDMD-NT protein was further purified using a Superdex 200 gel-filtration column. For protein used for bacterial growth assay, the Ni-NTA resin was washed with at least 40 column volumes of lysis buffer without detergent before elution and the protein was further purified with detergent-free buffers.

The caspase-11 gene was cloned into the pFastBac-HTa vector with a TEV cleavable N-terminal His_6-tag using EcoRI and XhoI restriction sites. The baculoviruses were prepared using the Bac-to-Bac system (Invitrogen), and the protein was expressed in Sf9 cells following the manufacturer's instructions. His–caspase-11 baculovirus (10 ml) was used to infect 1 l of Sf9 cells. Cells were collected 48 h after infection and His–caspase-11 was purified following the same protocol as for His_6-MBP–GSDMD. After elution from Ni-NTA resin, the protein was further purified using a Superdex 200 gel-filtration column. Aggregated fractions, which were the activated form of caspase-11, were collected for use in subsequent assays. For some experiments GSDMD-NT was generated by mixing caspase-11 with GSDMD at a 1:3 mass ratio for indicated times at 16 °C.

Native human granulysin and perforin were purified from isolated YT-Indy cytotoxic granules as previously described[30].

GSDMD thermal shift assay. Experimental protein unfolding was monitored by fluorescence of the Protein Thermal Shift Dye (Thermo Fisher Scientific) as temperature was continuously increased at a ramp rate of 1.6 °C per min using an Applied Biosystems StepOne Real-Time PCR machine. Samples of wild-type and 4A-mutant MBP–GSDMD were subdivided into three 20 μl replicates on a MicroAmp Optical 96-Well reaction plate. The transition thermal melting temperatures (T_m) were extracted using Applied Biosystems StepOne software version 2.3.

Immunoblot and immunoprecipitation. Cells were lysed in lysis buffer (50 mM Tris-Cl (pH 7.4), 150 mM NaCl) supplemented with 1% Triton-X100, 1 mM PMSF. Cell lysates were boiled in SDS loading buffer for 5 min before electrophoresis through SDS–PAGE gel. The resolved proteins were then transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore), which was probed with the indicated antibodies. Protein bands were visualized using a SuperSignal West Pico chemiluminescence ECL kit (Pierce). For non-reducing gels, cells were lysed in lysis buffer with or without 30 mM *N*-ethylmaleimide and cell lysates were prepared with 2ME-free SDS loading buffer. For immunoprecipitations, cell extracts were prepared using RIPA buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.5% deoxycholate) containing complete protease inhibitor cocktail. Lysates were incubated with the relevant antibody for 4 h at 4 °C before adding protein A/G agarose for 2 h. Beads were washed three times with the same buffer and bound proteins were eluted with SDS loading buffer by boiling for 5 min.

Native gel immunoblot. Cells samples, prepared using NativePAGE Sample Prep Kit (Invitrogen), were electrophoresed through a 4–16% NativePAGE Bis-Tris gel (Invitrogen) in NativePAGE running buffer (Invitrogen) at 4 °C and 150 V. Proteins were then transferred to a PVDF membrane at 0.2 A for 1 h in NativePAGE transfer buffer (Invitrogen) for immunoblotting.

Cytotoxicity assays. Cell death and cell viability were determined by the lactate dehydrogenase release assay using CytoTox 96 Non-Radioactive Cytotoxicity Assay kit (Promega) and by measuring ATP levels using the CellTiter-Glo Luminescent Cell Viability Assay (Promega), respectively, according to the manufacturer's instructions.

Triton X-114 phase separation. Cells were lysed in lysis buffer (20 mM HEPES (pH 7.4), 150 mM NaCl, 2% Triton X-114 (Sigma), complete protease inhibitor) and then centrifuged at 15,000*g* for 15 min. The resultant supernatant mixture was incubated at 30 °C for 10 min to separate the upper aqueous fraction from the lower detergent soluble fraction. The aqueous fraction was spun at 1,500*g* for 5 min at room temperature and the upper fraction harvested to eliminate contamination from the detergent-enriched phase. The detergent-enriched phase was diluted with lysis buffer lacking Triton X-114 and re-spun at 1,500*g* for 10 min and the detergent phase was recollected. The washed detergent phase was diluted with lysis buffer lacking Triton X-114 to the same final volume as the aqueous faction.

Cell fractionation. Cells were washed with PBS and collected by scraping in PBS on ice. Then cells were washed once in PBS and resuspended in five cell volumes of buffer A (20 mM HEPES (pH 7.4), 40 mM KCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 1 mM EGTA, 0.1 mM PMSF and 250 mM sucrose, 1× protease inhibitors). Cells were then incubated for 30 min on ice in buffer A, and lysed by passage through a 22-gauge needle 30 times. Lysates were spun at 800*g* for 10 min to remove unbroken cells and nuclei. The post-nuclear supernatant was spun at 7,000*g* for 10 min, and the supernatant (S7) was re–spun at 20,000*g* for 10 min, whereas the pellet (P7) was resuspended in the same volume of buffer A. The resulting pellet (P20) was again resuspended in buffer A and the supernatant (S20) was re-spun at 100,000*g* for 1 h and the resulting supernatant (S100) was collected and the pellet (P100) was

© 2016 Macmillan Publishers Limited. All rights reserved

resuspended in the same volume of buffer A as before. All the pellets containing membrane proteins were washed with buffer A. To separate soluble and crude membrane fractions, cells were lysed in buffer A and intact cells, nuclei and cell debris were removed by centrifugation of the homogenate at 800*g* for 10 min at 4 °C and then the supernatant was spun at 100,000*g* for 1 h at 4 °C. The supernatant containing cytosolic proteins (soluble fraction) was collected. The pellets were washed with buffer A and were re-centrifuged at 100,000*g* for 1 h. The precipitate was the crude membrane fraction.

Immunostaining and confocal microscopy. Cells grown on coverslips were fixed for 15 min with 4% paraformaldehyde in PBS, permeabilized for 20 min in 0.1% Triton X-100 in PBS and blocked using 5% BSA for 1 h. Then, the cells were stained with the indicated primary antibodies, followed by incubation with fluorescent-conjugated goat anti-mouse IgG (Invitrogen). Nuclei were counterstained with DAPI (Cell Signaling). Slides were mounted using Fluorescence Mounting Medium (Dako). Images were captured at room temperature using a confocal microscope (Olympus Fluoview FV1000 Confocal System) with a 63× water immersion objective and Olympus Fluoview software (Olympus). All confocal images are representative of three independent experiments.

Protein–lipid binding assay. Proteins were spotted on Membrane Lipid Strips (Echelon Biosciences) according to the manufacturer's instructions. To block non-specific binding, lipid strips were preincubated with binding assay buffer (3% fatty acid-free BSA (Sigma) in PBS) for 1 h at room temperature. Then the strips were incubated with protein (2 μg ml^{-1}) diluted in binding assay buffer for 1 h at room temperature and then washed three times (6 min each time) with wash buffer (0.1% Tween-20 in PBS). Membrane-bound proteins were detected by probing the lipid strips with corresponding primary antibodies diluted in binding assay buffer for 1 h at room temperature, followed by incubation for 1 h with horseradish-peroxidase-conjugated secondary antibody diluted 1:2000 in binding assay buffer. After washing three times with wash buffer, proteins were visualized using a SuperSignal West Pico chemiluminescence ECL kit (Pierce).

Liposome binding assay. Liposomes were prepared by hydration of lipids (Avanti Polar Lipids) in buffer R (20 mM HEPES (pH 7.4), 150 mM NaCl) followed by extrusion through a 100-nm polycarbonate membrane (~24 passages). All liposomes were composed of a 4:1 molar ratio of distearoyl PC and PE, to which we added other phospholipids (PtdIns(4)P, PtdIns(4,5)P2, PS, PA) as indicated. For liposome binding assay, protein (0.1 μM) was incubated with liposomes in 100 μl of buffer R for 20 min at room temperature before sedimentation at 140,000*g* for 20 min at 4 °C. Supernatants were removed immediately and the pellets were washed twice with buffer R and then resuspended in an equal volume of buffer. Proteins in both pellets and supernatant were then analysed by SDS–PAGE and immunoblot.

Liposome leakage assay. The leakage of liposomes encapsulating $TbCl_3$ was determined by an increase in fluorescence intensity when Tb^{3+} bound to DPA in the external buffer. Tb^{3+}-entrapped liposomes were prepared by hydration of the indicated lipids in buffer R containing 50 mM sodium citrate and 15 mM $TbCl_3$. Liposomes were washed twice to remove unincorporated $TbCl_3$. Then, Tb^{3+} entrapped liposomes were suspended in 100 μl buffer R supplemented with 50 μM of DPA and indicated GSDMD recombinant proteins were added. Fluorescence at 490 nm after excitation at 276 nm was continuously recorded for 9 min at 20 s intervals using a Biotek Synergy plate reader. At the end of the incubation, 0.1% Triton X-100 was added to the medium to measure complete release of Tb^{3+}. The extent of liposome leakage was calculated by using the formula R: $t\,(\%) = 100 \times ((F_t - F_{t0})/(F_{t100} - F_{t0}))$, where F_{t0} is the initial fluorescence of the Tb^{3+} liposomes in the DPA-containing buffer at the time GSDMD recombinant proteins were added, F_t is the fluorescence signal recorded at individual time points, and F_{t100} is the mean of the top-three fluorescence reads after adding 0.1% Triton X-100.

Preparation of unilamellar liposomes for reconstitution of GSDMD-NT pores. Synthetic 1,2-dioleoyl-*sn*-glycero-3-(phospho-L-serine) (DOPS), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) and 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE) (Avanti Polar Lipids) dissolved in chloroform were mixed in a glass tube at a mass ratio of 5:5:1, and the solvent was evaporated under a stream of N_2 gas. A buffer composed of 25 mM Tris-HCl (pH 8.0) and 150 mM NaCl was added to yield a final lipid concentration of 5 mM. The lipid suspension was then vortexed continuously for 5 min. To obtain unilamellar vesicles, liposomes were extruded with 21 passes through a mini-extruder device (Avanti) using membranes with 100 nm pores.

Reconstitution of GSDMD-NT pores on PS-containing liposomes. PS-containing liposomes (2 μmol) were incubated with 1 mg full length GSDMD and 0.3 mg caspase-11 for 4 h at 16 °C. After incubation, the liposome–protein suspension was collected by ultracentrifugation at 60,000*g* for 30 min at 4 °C. The pellets were washed twice with lysis buffer and then resuspended in 500 μl lysis buffer containing 0.5% C12E8 (Anatrace). After centrifugation for 30 min at 60,000*g*, the supernatant containing the solubilized GSDMD-NT pores was loaded in running buffer containing 0.5% C12E8 onto a Sepharose-6 gel filtration column. Fractions were analysed by SDS–PAGE and the fractions containing GSDMD-NT were pooled and imaged using negative staining electron microscopy.

Negative staining electron microscopy of GSDMD-NT pores. Copper grids (Electron Microscopy Sciences) coated with a layer of thin carbon were rendered hydrophilic immediately before use by glow-discharge in air with 25 mA current for 1 min. Liposome–protein suspensions (5 μl), or protein samples extracted from liposomes were loaded onto the grids, air dried for ~1 min and blotted, leaving a thin layer of sample on the grid surface. The grids were floated on a drop of staining solution containing 2% uranyl acetate for 60 s. After air drying, the grids were examined using a Tecnai G^2 Spirit BioTWIN electron microscope.

Bacterial growth assay. Colony-forming unit assays and turbidimetry were used to measure bacterial growth as previously described[11]. Briefly, for turbidimetry, bacteria were diluted (1:100) in bacterial culture medium following treatment and incubated with discontinuous shaking at 37 °C in a 200 μl volume in flat-bottomed 96-well plates. Growth curves were monitored by reading absorbance at 600 nm over 16 h using a Spectra MAX 340 (Molecular Devices) or Synergy H4 Hybrid Multi-Mode Microplate Reader (BioTek). The time until the growth curves reached a threshold OD_{600} of 0.05 above background was defined as the $T_{threshold}$. The ratio of $T_{threshold}$ (untreated):$T_{threshold}$ (treated) was used to quantify the change in bacterial growth.

LIVE/DEAD assay. Bacterial viability was assessed using the bacterial LIVE/DEAD assay (Invitrogen), following the manufacturer's recommendations. Briefly, bacteria were treated in the presence of 5 μM Syto-9 (Invitrogen) and 15 μM propidium iodide (Invitrogen). Treatment with 70% isopropanol served as a positive control. Fluorescence was visualized by confocal microscopy.

Fluorescent protein labelling and protein binding assay. Full-length and C-terminal GSDMD was labelled with AlexaFluor-488 using the Molecular Probes protein labelling kit. An aliquot of the labelled full-length protein was activated by incubating with active caspase-11 for 15 min at 37 °C. *L. monocytogenes* expressing mCherry (a gift from J. Theriot, Stanford Medical School) were treated with PBS or with 500 nM AF488-labelled GSDMD that had been activated or not with caspase-11 or with GSDMD-CT for 30 min at 37 °C. Bacteria were washed with 10 mM arginine in PBS for 10 min before fixation in 2% formalin in PBS. Slides were mounted with fluorescence mounting medium (Dako) and imaged using a fully motorized Axio Observer spinning disk microscope (Carl Zeiss Microimaging, Inc.) equipped with a cooled electron multiplication CCD camera with 512 × 512 resolution (Photometrics QuantEM, Tuscon, AZ) with excitation filters set at 405, 488, 561 and 640 nm and emission filter ranges of 430–475, 500–550, 589–625 and 680 nm long-pass, respectively. Images were analysed using SlideBook V5.0 (Intelligent Imaging Inc.) software. Third-dimensional image stacks were obtained along the *z* axis using the 63× oil immersion objective by acquiring sequential optical planes spaced 0.25 μm apart. Raw images were deconvolved using SlideBook.

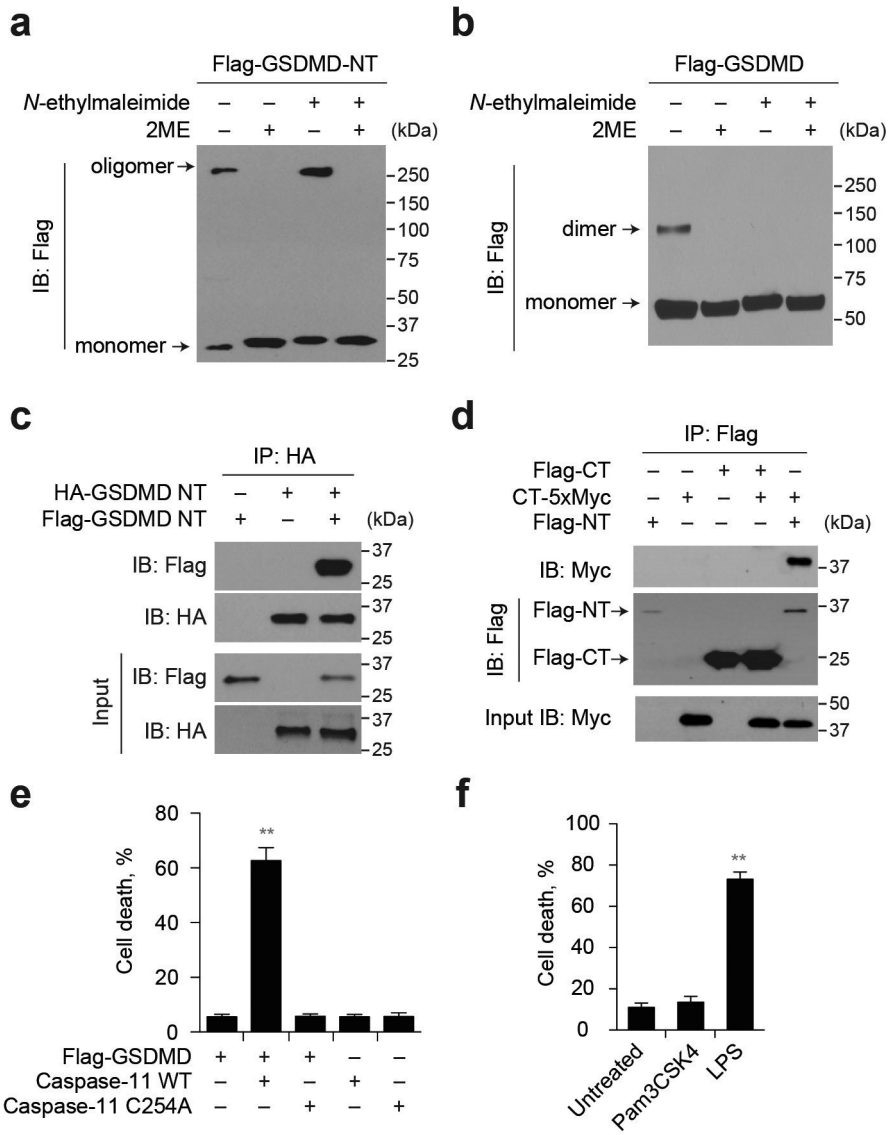
Treatment of extracellular bacteria. HEK293T cells, cultured in antibiotic-free medium, were transfected with the indicated plasmids for 30 h before culture supernatants were collected and concentrated fivefold. iBMDMs, cultured in antibiotic-free medium, were transfected with LPS or Pam3CSK4 or incubated with LPS and nigericin for 3 h before culture supernatants were collected and used without concentration. Exponential phase bacteria were treated with the indicated antibiotic-free culture supernatants or recombinant proteins, which were cultured at 37 °C for the indicated time. Treated bacteria were diluted in LB and plated on LB (*E. coli*, *S. aureus*) or brain–heart infusion agar (*L. monocytogenes*) agar plates to determine c.f.u., which were normalized to c.f.u. in control conditions.

Intracellular bacterial killing assay. HeLa cells or iBMDMs were transfected with indicated plasmids or siRNAs. Cells were infected with *L. monocytogenes* (multiplicity of infection, 10:1) 6 h after transfection of plasmids or 48 h after siRNA transfection. Cell plates were centrifuged at 1500 r.p.m. for 10 min, and placed at 37 °C for 30 min before washing to remove extracellular bacteria. Cells were lysed using 0.1% Triton-X100 at indicated time points after infection and supernatants were collected to determine bacterial titers by c.f.u. assay. For the nigericin experiment, iBMDMs were primed for 4 h with LPS (100 ng ml^{-1}) and then infected with *L. monocytogenes* as described above. After removing extracellular bacteria by washing, cells were treated or not with nigericin (20 μM) and 1 h later bacteria were collected as described above for c.f.u. assay.

Statistics. Student's *t*-test (two-tailed) was used for the statistical analysis of all experiments. *P* values <0.05 were considered significant.

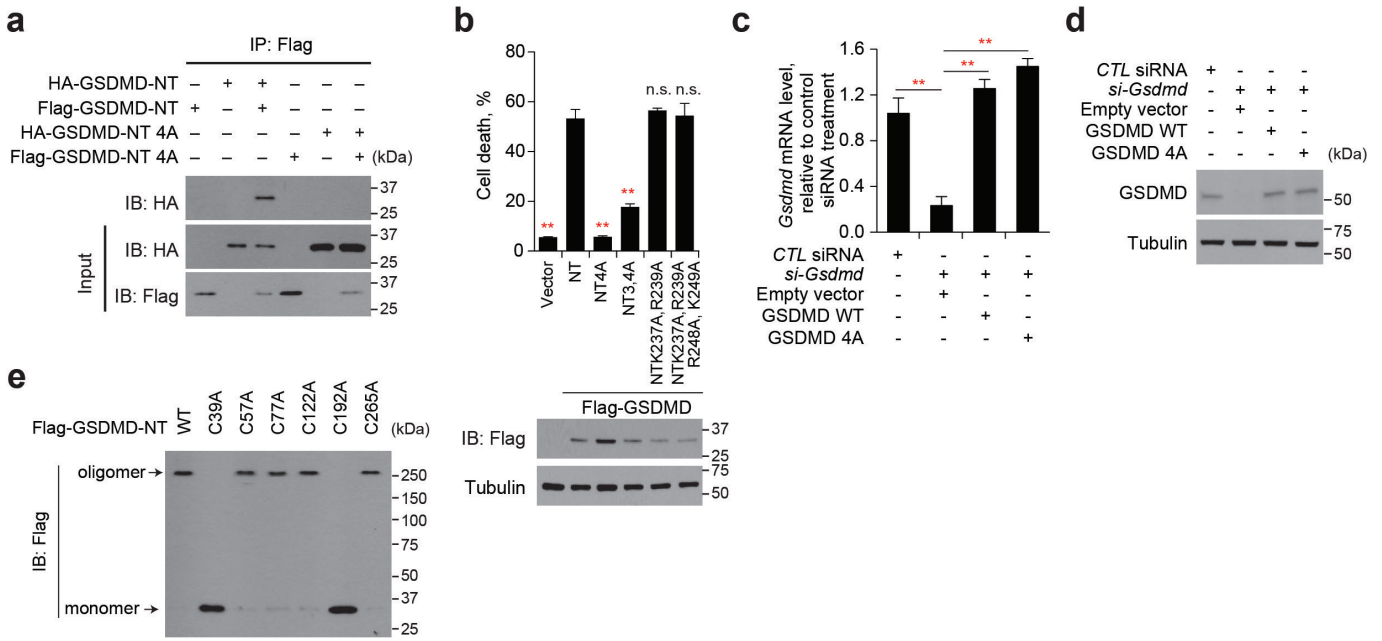
30. Thiery, J., Walch, M., Jensen, D. K., Martinvalet, D. & Lieberman, J. Isolation of cytotoxic T cell and NK granules and purification of their effector proteins. *Curr. Protoc. Cell Biol.* **47,** 3.37:3.37.1–3.37.29 (2010).

© 2016 Macmillan Publishers Limited. All rights reserved



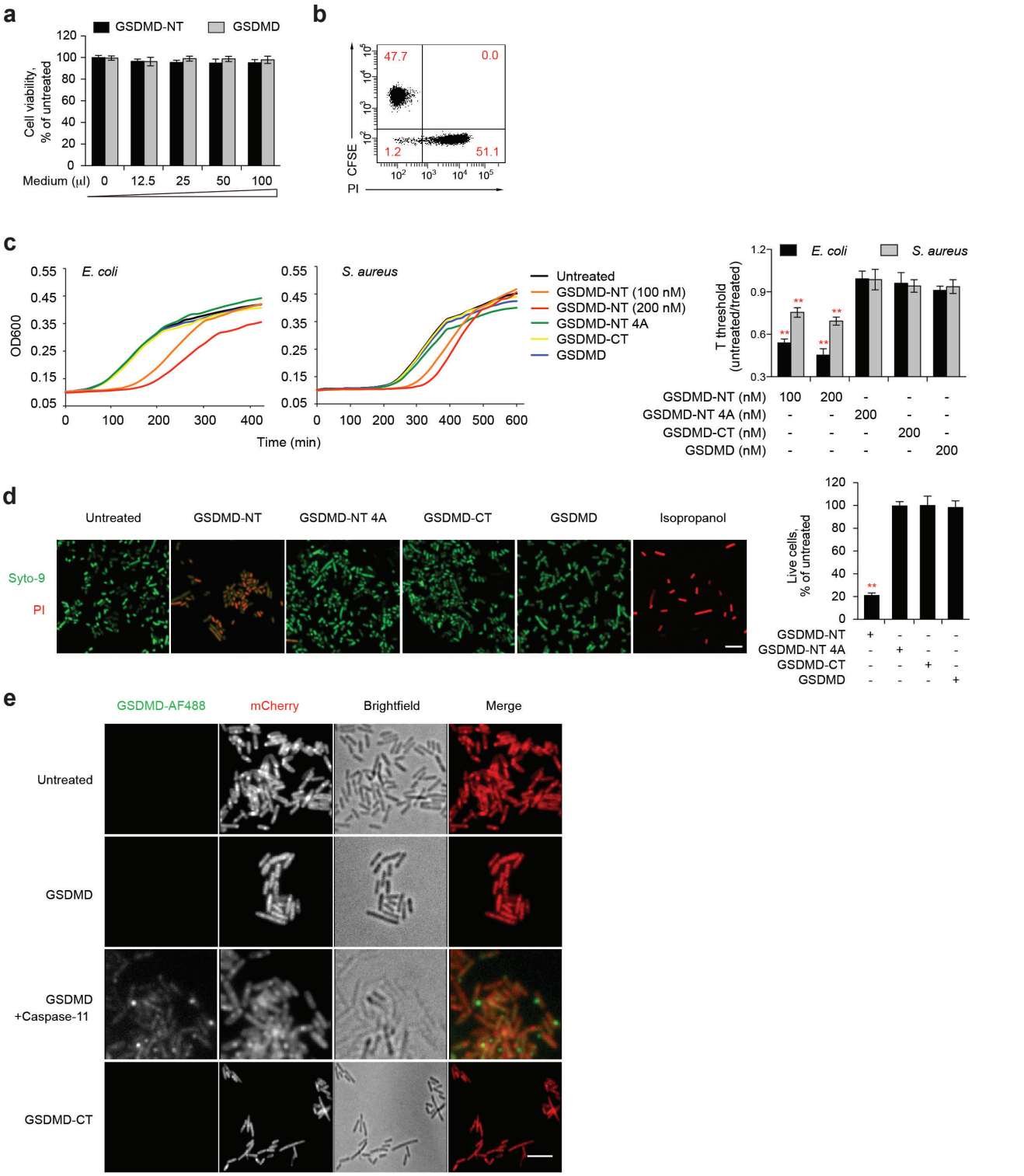
Extended Data Figure 1 | GSDMD-NT oligomerizes and induces pyroptosis. a, **b**, HEK293T cells, transfected with Flag–GSDMD-NT (**a**) or Flag–GSDMD (**b**), were lysed with or without *N*-ethylmaleimide or 2ME, and analysed by SDS–PAGE and Flag immunoblot. **c**, Lysates of HEK293T cells, transfected with HA–GSDMD-NT and/or Flag–GSDMD-NT, were immunoprecipitated with anti-HA and analysed by immunoblot with the indicated antibodies. **d**, HEK293T cells were transfected with the indicated plasmids. Cell lysates were immunoprecipitated with anti-Flag and analysed by immunoblot with the indicated antibodies. Flag–GSDMD-NT (Flag-NT) was expressed at considerably lower levels than GSDMD-CT-MYC (CT-MYC) or Flag–GSDMD-CT (Flag-CT), which accounts for the relative weak intensity of the corresponding bands on the middle blot. **e**, HEK293T cells, transiently transfected with the indicated plasmids, were assessed 16 h after transfection for cell death by CytoTox96 assay. **f**, Immortalized iBMDMs expressing Flag–GSDMD were electroporated with PBS, ultra LPS or Pam3CSK4, as a negative control for pyroptosis. 2 h later, cell death was determined by CytoTox96 assay. Graphs show the mean ± s.d. of triplicate wells and data shown are representative of three independent experiments. $^{**}P < 0.01$ (two-tailed *t*-test).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 2 | Mutation of four positively charged residues in GSDMD-NT or of two cysteine residues disrupts pyroptosis. **a**, Lysates of HEK293T cells, transfected with the indicated plasmids, were immunoprecipitated with anti-Flag and analysed by immunoblot with the indicated antibodies. The 4A mutant of GSDMD-NT does not self-associate in multimers. **b**, Mutations in other basic residues do not affect pyroptosis. The indicated wild-type or mutated Flag–GSDMD-NT constructs were transiently expressed in HEK293T cells. Medium was collected 18 h after transfection and cell death was measured by CytoTox96 assay. **c**, **d**, Knockdown in immortalized iBMDMs of *Gsdmd* and ectopic expression of wild-type or 4A *Gsdmd* mRNA (**c**, assessed by qRT–PCR relative to *GAPDH*) and protein (**d**, relative to tubulin). These data for the cells used in the rescue experiment in Fig. 1h show that the ectopic proteins are expressed at similar levels as the endogenous protein. **e**, Replacement of Cys37 or Cys192 by Ala in GSDMD-NT disrupts oligomerization. Mean ± s. d. of three technical replicates and data shown are representative of three independent experiments (**b**, **c**). Statistical differences are calculated by two-tailed *t*-test (in **b**, compared to samples transfected to express wild-type GSDMD-NT); $^{**}P < 0.01$ (two-tailed *t*-test).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 3 | Treatment with GSDMD-NT reduces bacterial viability, but does not affect the viability of mammalian cells. **a**, Antibiotic-free culture supernatants (concentrated fivefold) from transfected HEK293T cells, collected 30 h after transfection, were added to iBMDMs, which were cultured at 37 °C in 200 μl final volume for 6 h before measuring viability by CellTiter-Glo. **b**, HEK293T cells, transfected with Flag–GSDMD-NT 6 h earlier, were mixed with an equal number of CFSE-labelled untransfected HEK293T cells and incubated for 18 h before assessing cell death by propidium iodide staining and flow cytometry. **c**, *E. coli* and *S. aureus* were untreated or treated with recombinant GSDMD, wild-type or 4A-mutant GSDMD-NT, or GSDMD-CT (200 nM or indicated concentrations) for 20 min before samples were collected and bacterial growth was assessed by monitoring turbidity by optical density (representative experiments, left). The time to reach OD_{600} of 0.05 above background, which is a quantitative measure of the lag in detectable growth because of fewer viable bacteria, was defined as $T_{threshold}$ (right). The right graph shows the mean ± s.d. of three technical replicates. **d**, Bacterial viability after 20 min incubation with indicated proteins (200 nM) or isopropanol. Syto-9 enters live and dead bacteria, PI only enters dead bacteria (representative images, left; percent live cells, right). **e**, Fluorescence microscopy of mCherry-expressing *L. monocytogenes* incubated with AlexaFluor 488-GSDMD (activated or not with caspase-11) or AlexaFluor488-GSDMD-CT for 30 min at 37 °C. Data shown are representative of results of three independent experiments. Statistical differences are relative to untreated samples; $^{**}P < 0.01$ (two-tailed *t*-test). Scale bars, 5 μm.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18631

Genetic dissection of *Flaviviridae* host factors through genome–scale CRISPR screens

Caleb D. Marceau[1]*, Andreas S. Puschnik[1]*, Karim Majzoub[1], Yaw Shin Ooi[1], Susan M. Brewer[1], Gabriele Fuchs[1], Kavya Swaminathan[2], Miguel A. Mata[1], Joshua E. Elias[2], Peter Sarnow[1] & Jan E. Carette[1]

The *Flaviviridae* are a family of viruses that cause severe human diseases. For example, dengue virus (DENV) is a rapidly emerging pathogen causing an estimated 100 million symptomatic infections annually worldwide[1]. No approved antivirals are available to date and clinical trials with a tetravalent dengue vaccine showed disappointingly low protection rates[2]. Hepatitis C virus (HCV) also remains a major medical problem, with 160 million chronically infected patients worldwide and only expensive treatments available[3]. Despite distinct differences in their pathogenesis and modes of transmission, the two viruses share common replication strategies[4]. A detailed understanding of the host functions that determine viral infection is lacking. Here we use a pooled CRISPR genetic screening strategy[5,6] to comprehensively dissect host factors required for these two highly important *Flaviviridae* members. For DENV, we identified endoplasmic-reticulum (ER)-associated multi-protein complexes involved in signal sequence recognition, *N*-linked glycosylation and ER-associated degradation. DENV replication was nearly completely abrogated in cells deficient in the oligosaccharyltransferase (OST) complex. Mechanistic studies pinpointed viral RNA replication and not entry or translation as the crucial step requiring the OST complex. Moreover, we show that viral non-structural proteins bind to the OST complex. The identified ER-associated protein complexes were also important for infection by other mosquito-borne flaviviruses including Zika virus, an emerging pathogen causing severe birth defects[7]. By contrast, the most significant genes identified in the HCV screen were distinct and included viral receptors, RNA-binding proteins and enzymes involved in metabolism. We found an unexpected link between intracellular flavin adenine dinucleotide (FAD) levels and HCV replication. This study shows notable divergence in host-depenency factors between DENV and HCV, and illuminates new host targets for antiviral therapy.

CRISPR is revolutionizing the use of genetic screens because the ability to completely knockout genes substantially increases the robustness of the phenotypes[5,6]. We compared the CRISPR approach in the hepatocyte cell line Huh7.5.1 with an alternative method to generate knockout alleles on a genome-wide scale: insertional mutagenesis in human haploid cells (HAP1)[8,9] (Fig. 1a). Both methods generate libraries of cells with knockout mutations in all non-essential genes. To comprehensively identify cellular genes with crucial roles in the *Flaviviridae* life cycles, we first infected pools of mutagenized cells with DENV serotype 2 (DENV-2). The two types of genetic screening methods showed a strong concordance in the genes enriched in the DENV-2-resistant population. Many could be functionally classified into three distinct categories, each important for proper expression of ER-targeted glycoproteins (Fig. 1b, c, Supplementary Tables 1, 2). The translocon associated protein (TRAP) complex (containing subunits SSR1, SSR2 and SSR3) has an elusive role in stimulating co-translational translocation mediated by several, but not all, signal sequences[10] (Fig. 1b, c, blue). Genes involved in protein quality control and the ER-associated protein degradation (ERAD) pathway also scored highly (Fig. 1b, c, green). Notably, in both the haploid and CRISPR genetic screens, the most significantly enriched genes were subunits of the OST complex, an enzyme essential for *N*-linked glycosylation (Fig. 1b, c, red). This dependence on ER cellular genes is probably related to the expression of the DENV genome, which encodes an ER-targeted viral polyprotein containing signal sequences and viral glycoproteins. Given the similarities in DENV and HCV polyprotein expression, we expected these genes to also be represented in the HCV CRISPR screen. Surprisingly, there was no overlap between the DENV and HCV core sets of enriched genes, suggesting that these members of the *Flaviviridae* evolved divergent host factor dependencies (Fig. 1c–e, Extended Data Fig. 1a, b, Supplementary Tables 3, 4). Indeed, cross-comparison of the most significant hits with both viruses suggested specific dependencies, although minor quantitative effects cannot be excluded (Extended Data Fig. 1c). The robustness of the CRISPR approach was further underscored by the consistent identification of the core dependency factors in three independent replicate screens performed for each virus (Extended Data Fig. 2). We validated the novel DENV host factors in isogenic knockout cells using a plaque-forming assay and observed a marked reduction in particle formation (Extended Data Figs 3, 4a). Importantly, complementation of knockout cells restored DENV infection (Extended Data Fig. 4b, c). The relevance of the identified host factors was further confirmed in Raji DC-SIGN, a B-cell line commonly used to study DENV (Extended Data Fig. 4d).

Struck by the distinct host factor requirements of DENV-2 and HCV, we sought to evaluate selected DENV-2-dependency factors against other mosquito-borne flaviviruses that are closely related to DENV (Fig. 2a). Using quantitative PCR (qPCR) in isogenic knockout cells, we found that West Nile virus (WNV), but not yellow fever virus (YFV) or Zika virus (ZIKV), was as sensitive as DENV-2 to the disruption of the tested ERAD genes, which is in line with previous reports implicating ERAD in WNV infection[11,12]. A functional TRAP complex is important for DENV-2, YFV and ZIKV RNA replication, whereas WNV RNA abundance is only slightly reduced. Individual subunits of the OST complex displayed notably different phenotypes for the four related flaviviruses. Whereas knockout of *STT3A* and *STT3B* both completely abolished DENV-2 replication, only *STT3A* knockout affected YFV, WNV and ZIKV replication. When probing HCV replication in *STT3A*- and *STT3B*-knockout Huh7 cells using luciferase virus, we did not observe a substantial decrease (Extended Data Fig. 4e).

Intrigued by the differential sensitivity to the catalytic OST subunits, we focused our mechanistic studies on the OST complex, which has not been linked to viral replication before. The highly conserved catalytic subunit of the OST complex, STT3, is duplicated into two paralogues STT3A and STT3B in mammalian cells, and each isoform is present

[1]Stanford University, Department of Microbiology and Immunology, Stanford, California 94305, USA. [2]Stanford University, Department of Chemical and Systems Biology, Stanford, California 94305, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved

NOT FINAL PROOF

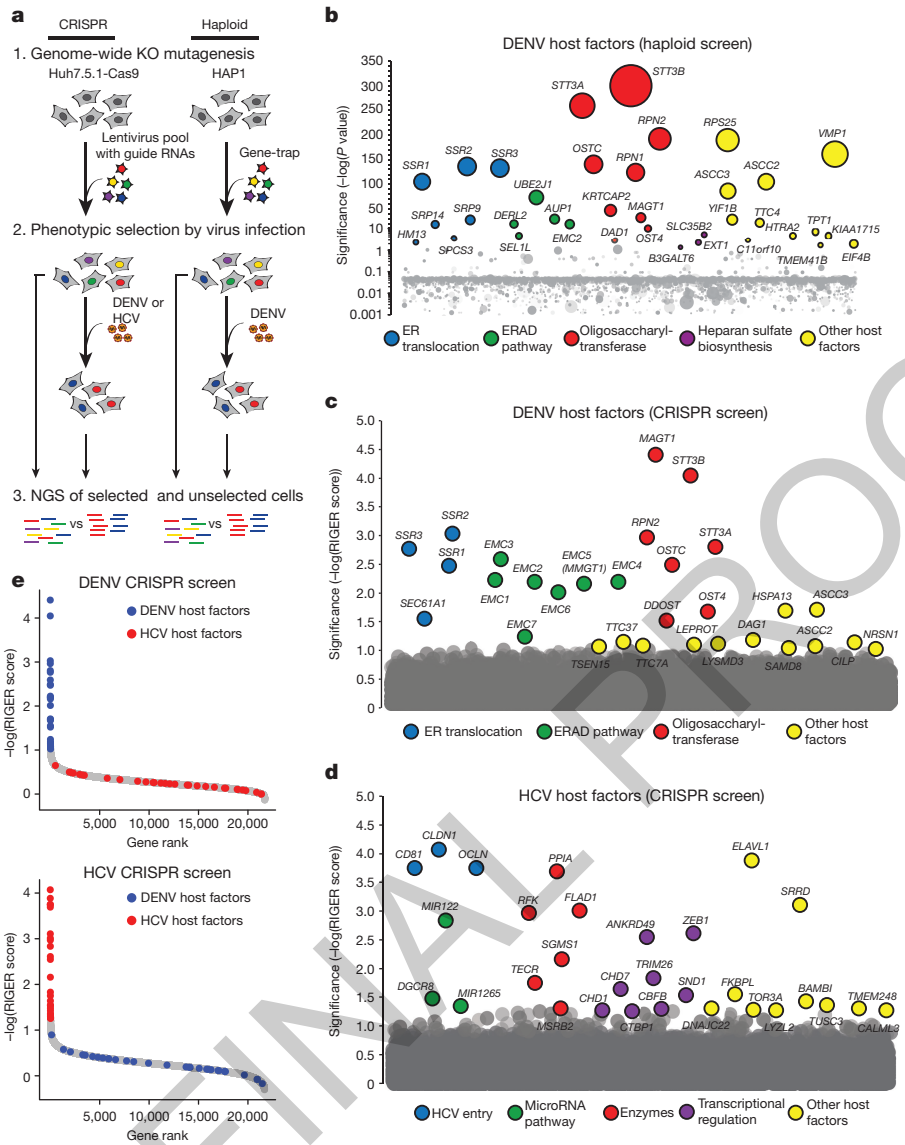


Figure 1 | Haploid and CRISPR genetic screens identify essential host factors of DENV and HCV infections. a, Schematic for genome-wide screening approach. NGS, next-generation sequencing. **b**, Haploid genetic screen for DENV host factors. The *y* axis represents significance of enrichment of gene-trap insertions in genes in DENV-resistant population compared to unselected HAP1 cells. Each circle represents a specific gene and size corresponds to the number of independent gene-trap insertions. All genes with $P < 0.05$ (Fisher's exact test) were coloured and grouped by function. The screen was performed once. **c**, **d**, CRISPR genetic screen for DENV (**c**) and HCV (**d**) host factors in Huh7.5.1 cells. Significance of enrichment was calculated by RIGER analysis. The screens were performed in three replicates and the mean of the RIGER score is represented on the *y* axis. The 30 most enriched genes were coloured and grouped by function. **e**, Comparison of the 30 most enriched genes from the DENV and HCV CRISPR screens and their position based on the mean RIGER score.

in distinct protein complexes[13] (Extended Data Fig. 4f). The STT3A complex is important for the co-translational *N*-linked glycosylation of most of the glycoproteins, whereas the STT3B complex is important for the co-translational or post-translational glycosylation of acceptor sites that have been skipped by the STT3A complex[14]. Despite their partially redundant functions in *N*-linked glycosylation, we found that all DENV serotypes required the presence of both catalytic subunits as well as MAGT1, the highest scoring gene in the CRISPR screen (Fig. 2b).

To pinpoint which step in the viral life cycle requires the OST complex, we first focused on viral entry. We did not observe major differences in viral particle entry in OST-deficient cells (Fig. 2c). Next, we used a replicon assay that bypasses viral entry by electroporation of DENV RNA. Translation of the viral genome, apparent at time points up to 10 h, was equally efficient in OST-knockout cells as in wild-type cells (Fig. 2d). In stark contrast, viral RNA replication (apparent at time points after 10 h) was completely abolished. This mirrored the expression pattern observed with a replication-deficient dengue mutant in the viral polymerase ($NS5^{GDD}$). Thus, we show that the OST complex has a crucial involvement in viral RNA replication, after entry and translation of the viral genome.

Most glycoproteins can be efficiently modified by both OST isoforms and only a few are preferentially modified by either STT3A or STT3B[14]. Concordant with this, knockout of either *STT3A* or *STT3B* did not lead to loss of cellular viability, whereas a double knockout was lethal (Fig. 3a). We demonstrated that OST catalytic activity is required for cellular function using STT3A and STT3B mutants containing mutations in the residues that coordinate the binding of the divalent cation required for catalysis[15] (Fig. 3a, Extended Data Fig. 5). The functional redundancy between isoforms in global *N*-linked glycosylation is in contrast with the extreme dependence on each individual isoforms of the OST complex for DENV replication. To investigate whether the effect of the OST complex is meditated by the necessity to glycosylate

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

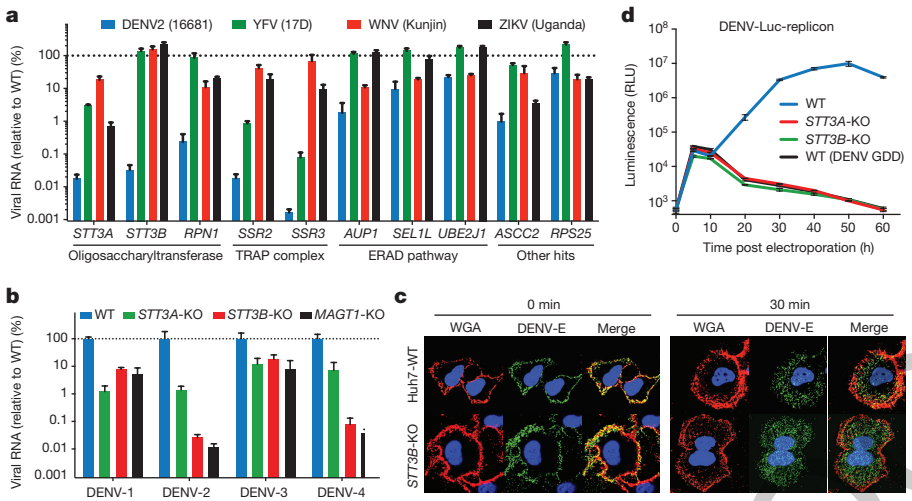


Figure 2 | ER protein complexes have a crucial role in the replication phase of DENV and are also important for YFV, WNV and ZIKV infection. a, qPCR of DENV (clone 16681), YFV (17D), WNV (Kunjin), and ZIKV (Uganda) RNA in knockout HAP1 cells. WT, wild type. **b**, qPCR of prototypic strains of DENV serotypes 1–4 RNA in knockout (KO) Huh7 cells. **c**, Confocal microscopy of *STT3B*-KO Huh7 cells immunostained for DENV envelope (DENV-E) protein immediately or 30 min after DENV infection. Original magnification, ×630. **d**, Luminescence of DENV replicon RNA expressing luciferase in knockout Huh7 cells. The DENV $NS5^{GDD}$ mutant served as replication-deficient control. RLU, relative light units. Data are mean and s.e.m. (qPCR) or s.d. (RLU) for triplicate infections.

viral proteins properly, we focused on NS1, an enigmatic DENV glycoprotein with essential roles in RNA replication[16] and pathogenesis[17]. NS1 was fully glycosylated in *STT3A*- and *STT3B*-knockout cells in contrast to the hypo-glycosylation of control cellular proteins known to be preferentially glycosylated by STT3A (pSAP) or STT3B (SHBG) (Fig. 3b). This led us to speculate that the OST complex itself rather than its catalytic activity is required for DENV replication. To explore this hypothesis, we used the catalytic mutants of STT3A and STT3B (Fig. 3a, c, Extended Data Fig. 5). Surprisingly, the catalytically dead mutants were able to restore fully DENV replication in *STT3A*- and in *STT3B*-knockout cells (Fig. 3d). We thus concluded that DENV RNA replication has hijacked a function of the human OST complex that is independent from its canonical role in *N*-linked glycosylation. The dispensability of the catalytic function of the OST complex further suggests a more structural role of OST in viral replication. The OST complex forms a stoichiometric complex at the ER membrane where DENV establishes a functional replication complex to initiate RNA replication. Our electron microscopy studies using APEX2 confirmed the localization of STT3B in the ER membrane, in close proximity to the membranous replication vesicles in DENV-2-infected cells (Extended Data Fig. 6a–c).

We next interrogated the physical interaction of the viral proteins with the OST complex by immunoprecipitation of STT3A–Flag and STTB–Flag in the context of viral infection. Western blot analysis after immunoprecipitation showed that the non-structural proteins NS2B and NS3, components of the DENV replication complex, associate specifically with STT3A and STT3B (Fig. 3e, Extended Data Fig. 6d). Next, eluates of the immunoprecipitations (Extended Data Fig. 6e, f) were subjected to tryptic proteolysis and the resulting peptides were analysed by mass spectrometry. Identification of tryptic peptides from NS2A, NS3 and NS4A in the eluates pointed to their association with the OST complexes formed by STT3A and STT3B (Extended Data Fig. 6g,

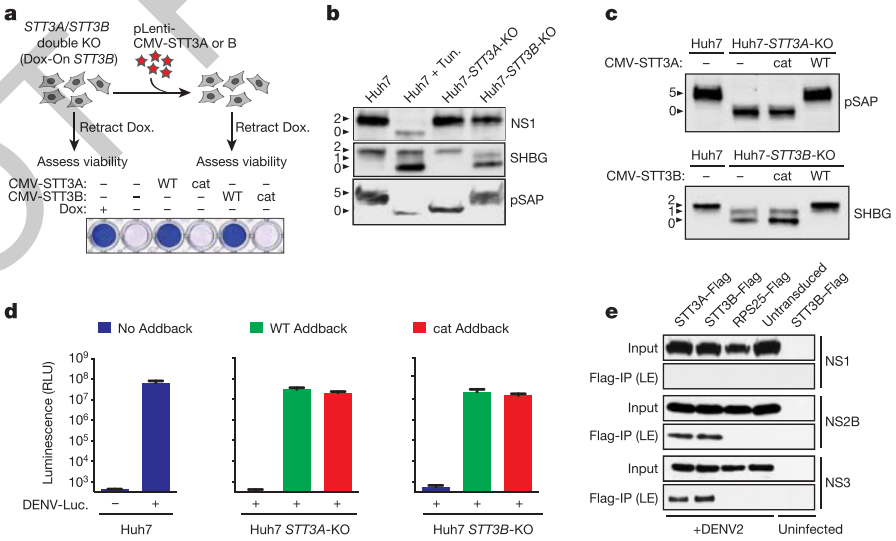


Figure 3 | DENV RNA replication requires a non-canonical function of OST, and DENV non-structural proteins interact with OST. a, Viability of *STT3A* and *STT3B* double-knockout cells complemented with wild-type (WT) or catalytic (cat) mutant cDNA. Dox, doxycycline. **b**, Glycosylation of DENV protein NS1, SHBG and pSAP in *STT3A*- and *STT3B*-knockout Huh7 cells. Different glycoforms are indicated by arrowheads. Tun., tunicamycin. **c**, Glycosylation state of pSAP and SHBG in *STT3A*- and *STT3B*-knockout cells complemented with catalytic mutants. **d**, DENV infection of knockout Huh7 cells complemented with wild-type or catalytic mutants of STT3A and STT3B. Data are mean and s.d. of triplicate infections. **e**, Co-immunoprecipitations of STT3A–Flag and STT3B–Flag from DENV-infected cell lysates. LE, long exposure.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

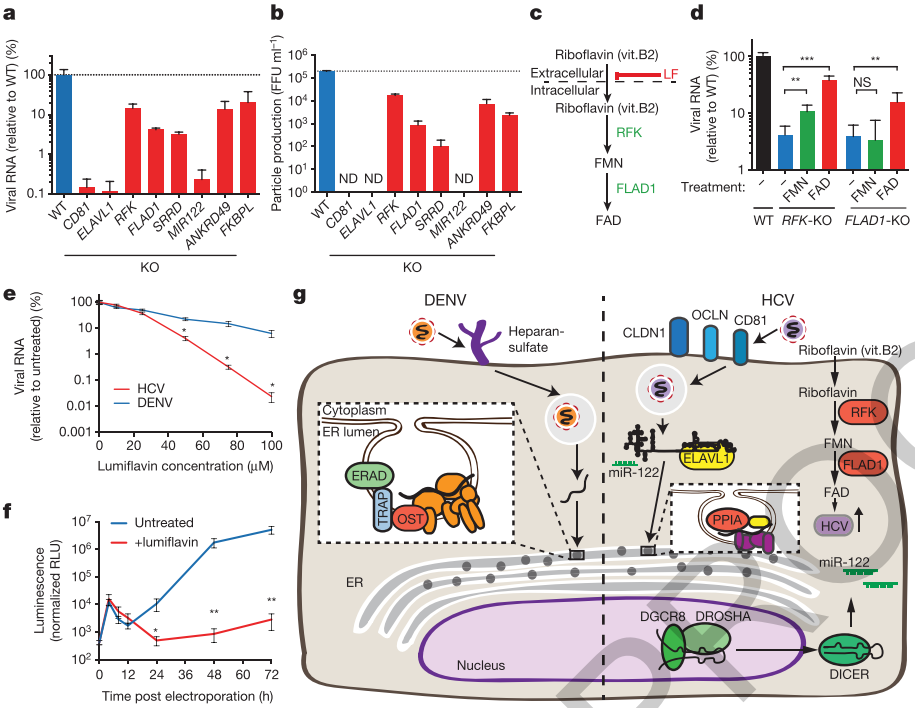


Figure 4 | FAD biosynthesis is required for HCV replication and can serve as antiviral target. a, qPCR of HCV RNA in Huh7.5.1 cell lines. **b**, HCV particle formation measured by focus-forming units (FFU) assay. ND, no foci detected (threshold of detection is 50 FFU ml^{-1}). **c**, Biosynthesis pathway of FAD. Lumiflavin (LF) competitively inhibits uptake of riboflavin. Vit.B2, vitamin B2. **d**, qPCR of HCV RNA in untreated, FMN- or FAD-treated *RFK*- and *FLAD1*-knockout Huh7.5.1 cells. **e**, qPCR of DENV or HCV RNA in lumiflavin-treated Huh7.5.1 cells. For each concentration, the significance of the effect on HCV versus DENV was determined. **f**, HCV replicon assay in untreated and lumiflavin-treated Huh7.5.1 cells using wild-type sgJFH1 replicon. **g**, Model of identified DENV and HCV host factors. Data are mean and s.e.m. (qPCR) or s.d. (FFU, RLU) for triplicate infections. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (unpaired, parametric two-sided Student's *t*-test, with Welch post-correction). NS, non-significant.

Supplementary Table 5). Taken together, this suggests a structural role for the OST complex in DENV RNA replication through interactions with non-structural proteins that form the RNA replication complex. Our data indicate that the OST complex fulfils specialized roles in host pathogen interactions and is more multifaceted than previously recognized. In support of this emerging notion, two recent studies uncovered unexpected roles of the OST complex in immunity. The OST complex was found to be crucial for innate immune responses triggered by lipopolysaccharide[18], and in a separate study OST dysregulation was identified as cause for autoimmune disorders triggered by certain TREX1 mutations[19].

The HCV-resistant cell population was highly enriched in guide RNAs targeting the known HCV receptors[20] *CD81*, *OCLN* and *CLDN1*, confirming their non-redundant role in entry for HCV (Fig. 1d, blue) and highlighting the validity of the screen results. The completeness of the screen was further underscored by the identification of microRNA-122 (ref. 21) and *DGCR8* (Fig. 1d, green), which is part of the microRNA processing machinery, as key factors for HCV replication. Several dependency factors of HCV were validated in Huh7.5.1 cells, where knockout significantly reduced viral RNA and particle formation (Fig. 4a, b, Extended Data Figs 7, 8a). After *CLDN1*, the second most significantly enriched gene was *ELAVL1* (also known as *HUR*), an RNA-binding protein involved in mRNA stabilization[22]. In isogenic *ELAVL1*-knockout cells, HCV RNA replication was nearly abolished, while we did not observe a decrease with other RNA viruses (Extended Data Fig. 8b, c) including the alphavirus Sindbis, which contains strong ELAVL1-binding sites[23]. We used an HCV replicon assay to show that ELAVL1 has a critical role in HCV RNA replication, which is in line with a recent report[24] (Extended Data Fig. 8d, e). Enzymatically active HCV-dependency factors (Fig. 1d, red) are by far the most important category of host factors for identifying antiviral drug targets. A case in point is cyclophilin A (PPIA) that has been actively pursued until phase III clinical trials[25]. We discovered enzymes with novel putative roles in HCV replication and explored these potentially 'druggable' factors further. We focused on RFK and FLAD1, enzymes involved in the two-step conversion of riboflavin (vitamin B2) to flavin adenine dinucleotide (FAD) (Fig. 4c). *RFK*- and *FLAD1*-knockout cells were resistant to HCV replication but not DENV (Extended Data Fig. 9a). As predicted from their sequential role in FAD biogenesis, exogenous flavin mononucleotide (FMN) or FAD rescued HCV replication in *RFK*-knockout cells, whereas FAD but not FMN rescued viral replication in *FLAD1*-knockout cells (Fig. 4d, Extended Data Fig. 9b). This demonstrates that HCV replication is dependent solely on sufficient FAD levels. Modulation of intracellular FAD levels can be achieved by treatment of the cells with lumiflavin, a cellular uptake inhibitor of riboflavin[26]. Treatment of cells with lumiflavin greatly reduced HCV replication, while other RNA viruses were less sensitive to lumiflavin treatment (Fig. 4e, Extended Data Fig. 9b–f). We further pinpointed RNA replication as the step of the life cycle that requires FAD using a replicon system (Fig. 4f). This highlights that knockout screens can identify specific host targets for antiviral drug discovery. Taken together, we used comparative genome-scale knockout screens to identify human genes with crucial roles in the replication of *Flaviviridae*. Despite previous extensive interrogation of human host factors for these viruses through genomic and proteomic approaches[11,27–30], we discovered marked dependencies on several host processes that had not been linked to flaviviral replication before (Fig. 4g, Extended Data Fig. 10).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January; accepted 10 June 2016.
Published online 17 June 2016.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

1. Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496,** 504–507 (2013).
2. Thomas, S. J. & Rothman, A. L. Trials and tribulations on the path to developing a dengue vaccine. *Am. J. Prev. Med.* **49** (suppl. 4), S334–S344 (2015).
3. Lavanchy, D. Evolving epidemiology of hepatitis C virus. *Clin. Microbiol. Infect.* **17,** 107–115 (2011).
4. Paul, D. & Bartenschlager, R. Flaviviridae replication organelles: oh, what a tangled web we weave. *Annu. Rev. Virol.* **2,** 289–310 (2015).
5. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–87 (2014).
6. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–84 (2014).
7. Rasmussen, S. A., Jamieson, D. J., Honein, M. A. & Petersen, L. R. Zika virus and birth defects—reviewing the evidence for causality. *N. Engl. J. Med.* **374,** 1981–1987 (2016).
8. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326,** 1231–1235 (2009).
9. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477,** 340–343 (2011).
10. Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* **160,** 529–539 (2003).
11. Krishnan, M. N. *et al.* RNA interference screen for human genes associated with West Nile virus infection. *Nature* **455,** 242–245 (2008).
12. Ma, H. *et al.* A CRISPR-based screen identifies genes essential for West-Nile-Virus-induced cell death. *Cell Rep.* **12,** 673–683 (2015).
13. Shrimal, S., Cherepanova, N. A. & Gilmore, R. Cotranslational and posttranslocational *N*-glycosylation of proteins in the endoplasmic reticulum. *Semin. Cell Dev. Biol.* **41,** 71–78 (2015).
14. Ruiz-Canada, C., Kelleher, D. J. & Gilmore, R. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell* **136,** 272–283 (2009).
15. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
16. Lindenbach, B. D. & Rice, C. M. trans-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. *J. Virol.* **71,** 9608–9617 (1997).
17. Beatty, P. R. *et al.* Dengue virus NS1 triggers endothelial permeability and vascular leak that is prevented by NS1 vaccination. *Sci. Transl. Med.* **7,** 304ra141 (2015).
18. Parnas, O. *et al.* A Genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* **162,** 675–686 (2015).
19. Hasan, M. *et al.* Cytosolic nuclease TREX1 regulates oligosaccharyltransferase activity independent of nuclease activity to suppress immune activation. *Immunity* **43,** 463–474 (2015).
20. Scheel, T. K. & Rice, C. M. Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. *Nat. Med.* **19,** 837–849 (2013).
21. Jopling, C. L., Yi, M., Lancaster, A. M., Lemon, S. M. & Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. *Science* **309,** 1577–1581 (2005).
22. Brennan, C. M. & Steitz, J. A. HuR and mRNA stability. *Cell. Mol. Life Sci.* **58,** 266–277 (2001).
23. Sokoloski, K. J. *et al.* Sindbis virus usurps the cellular HuR protein to stabilize its transcripts and promote productive infections in mammalian and mosquito cells. *Cell Host Microbe* **8,** 196–207 (2010).
24. Shwetha, S. *et al.* HuR displaces polypyrimidine tract binding protein to facilitate la binding to the 3′ untranslated region and enhances hepatitis C virus replication. *J. Virol.* **89,** 11356–11371 (2015).
25. Lin, K. & Gallay, P. Curing a viral infection by targeting the host: the example of cyclophilin inhibitors. *Antiviral Res.* **99,** 68–77 (2013).
26. Fujimura, M. *et al.* Functional characteristics of the human ortholog of riboflavin transporter 2 and riboflavin-responsive expression of its rat ortholog in the small intestine indicate its involvement in riboflavin absorption. *J. Nutr.* **140,** 1722–1727 (2010).
27. Sessions, O. M. *et al.* Discovery of insect and human dengue virus host factors. *Nature* **458,** 1047–1050 (2009).
28. Li, Q. *et al.* A genome-wide genetic screen for host factors required for hepatitis C virus propagation. *Proc. Natl Acad. Sci. USA* **106,** 16410–16415 (2009).
29. Tai, A. W. *et al.* A functional genomic screen identifies cellular cofactors of hepatitis C virus replication. *Cell Host Microbe* **5,** 298–307 (2009).
30. Ramage, H. R. *et al.* A combined proteomics/genomics approach links hepatitis C virus infection with nonsense-mediated mRNA decay. *Mol. Cell* **57,** 329–340 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank T. Brummelkamp, J. Idoyaga and S. Einav for critically reading the manuscript and discussions; X. Ji and the Stanford Functional Genomics Facility; The Stanford Transmission Electron Microscope Facility; The Stanford Shared FACS Facility, and members of the Carette laboratory for discussions and support. S. Einav, K. Kirkegaard, F. Chisari, J. F. Anderson, C. Rice, H. Ploegh, R. Kopito, E. Harris, M. Ivessa, S. Weaver, R. Tesh and E. Campeau are acknowledged for providing materials. The work was funded in part by National Institutes of Health (NIH) DP2 AI104557 (J.E.C.), NIH AI109662 (J.E.C.), David and Lucile Packard Foundation (J.E.C.), Stanford Graduate Fellowship (A.S.P.), Boehringer Ingelheim Fonds (A.S.P.) and NSF- GFRP (C.D.M.).

Author Contributions C.D.M., A.S.P. and J.E.C. were responsible for overall design of the study. Most of the experiments related to the DENV and HCV genetic screens were performed by C.D.M. and A.S.P., respectively. K.M. designed and performed several validation experiments and J.E.C. designed dengue constructs. Y.S.O. performed one of the DENV screens. P.S. provided expertise in design of HCV experiments and M.A.M. and S.M.B. performed several HCV experiments. Mass spectrometry experiments were performed and analysed by C.D.M., G.F. and K.S. under the technical expertise of J.E.E. C.D.M., A.S.P., K.M. and J.E.C. wrote the manuscript with input from all authors.

Author Information The CRISPR and haploid genetic screens have been deposited in the NCBI BioProject database under accession numbers PRJNA322191 and PRJNA284536, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E.C. (carette@stanford.edu).

Reviewer Information *Nature* thanks W. Wei and the other anonymous reviewer(s) for their contribution to the peer review of this work.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Haploid genetic screen. The haploid genetic screen was performed as previously described[9]. In short, 100 million gene-trap mutagenized HAP1 cells were seeded and infected with DENV-2 16681 (multiplicity of infection (MOI) = 5). Eight hours after infection, media was aspirated and replaced with IMDM containing 2% FBS. Clear cytopathic effects were observed after 3 days of infection leading to the death of most cells. Clusters of cells resistant to DENV-2 infection became apparent during further culture, and 9 days after infection cells were collected as a pool (yield ~30 million cells) and genomic DNA was isolated using a QIAamp DNA column. Gene-trap insertion sites were determined by linear amplification of the genomic DNA (gDNA) flanking regions of the gene-trap DNA insertion sites and sequenced on a Genome Analyzer II. Reads were aligned to the human genome using Bowtie and enrichment of independent insertions was calculated as previously described[9].

CRISPR genetic screens. Huh7.5.1 cells were stably transduced with lenti-Cas9-Blast and subsequently selected using blasticidin. Next, a total of 300 million Huh7.5.1 cells that constitutively express Cas9 were transduced with the lentiGuide-Puro from the GeCKO v2 library[31] at a MOI of 0.3. Cells were selected with puromycin and pooled together. The CRISPR genetic screens were started 10 days after transduction. Approximately 60 million mutagenized cells for each library (A and B) were infected with DENV-2 16681 (replicate 1 and 2) or DENV-2 strain 429557 (replicate 3) using a MOI of 1, or with HCV JFH1 at a MOI of 0.3. Cytopathic effect was visible 2 and 5 days after infection for DENV and HCV, respectively. Huh7.5.1 cells grow slower than HAP1 cells and clusters of resistant cells took longer to develop. The selected cells were collected 16 days after infection. As an uninfected reference we chose the unselected starting population because in these strong positive selection screens the selection pressure of the viral infections renders potential small growth differences of the mutagenized cells inconsequential. For both selected and uninfected control cells, gDNA was isolated using a QIAamp DNA column, and the inserted guide RNA sequences were amplified from the gDNA by flanking primers and prepared for next-generation sequencing. Resulting amplicons were sequenced on a MiSeq or NextSeq platform (Illumina) and the enrichment of each guide RNA was calculated by comparing the relative abundance in the selected and unselected population. RIGER analysis was performed on guide RNAs (with at least 10 reads) ranked by enrichment using the weighted sum statistical method[32]. Each CRISPR screen was performed in three replicates and the mean of the three RIGER scores was calculated.

Cell culture. HAP1 cells were derived from the near-haploid chronic myeloid leukaemia cell line KBM7 as described earlier[9]. HAP1 cells and knockout derivatives were cultured in IMDM supplemented with 10% FBS, penicillin-streptomycin and L-glutamine. *STT3A*/*STT3B* double-knockout HAP1 cells were cultured in IMDM supplemented with 10% FBS, penicillin-streptomycin, L-glutamine and 25 ng ml^{-1} doxycycline. Huh7, Huh7.5.1 (both gifts from F. Chisari) and HEK293FT (Thermo Scientific) cells and knockout derivatives were grown in DMEM supplemented with 10% FBS, penicillin-streptomycin, non-essential amino acids and L-glutamine. HEK293FT cells were used to generate lentivirus vectors for cellular transductions. Raji DC-SIGN cells (a gift from E. Harris) were cultured in RPMI-1640 supplemented with 10% FBS, penicillin-streptomycin and L-glutamine. The cell lines have not been authenticated. Parental cell lines have been tested negative for mycoplasma.

Viral serotypes and strains. DENV-2 infectious clone 16681 was a gift from K. Kirkegaard. DENV-2 from infectious clone 16681 was adapted to HAP1 cells through serial passaging. Viral whole-genome sequence analysis revealed three coding mutations compared to the original clone 16681: Q399H in envelope, L180F in NS2A and S238F in NS4B. DENV-1 Hawaii 1944 (#NR82), DENV-2 strain 429557 (#NR-12216), DENV-2 New Guinea C 1944 (#NR-84), DENV-3 Philippines/H871856 (#NR-80) and DENV-4 H241 Philippines 1956 (#NR-86) were ordered from BEI resources (NIH, NIAID). Yellow fever virus was generated by culturing the yellow fever YF-VAX 17D-204 vaccine. West Nile virus (Kunjin strain CH 16532) was a gift from J. F. Anderson. Zika virus (strain MR766) was provided by S. Weaver and R. Tesh. Hepatitis C virus JFH1 and HCV-Luc pFL-J6/JFH-5′C19Rluc2AUbi vector were gifts from C. Rice. Sindbis virus (SINV) strain Ar-339 (TC adapted) Egypt 1952 (ATCC VR-1585) and human rhinovirus 14 (ATCC VR-284) were ordered from the American Type Culture Collection. Poliovirus type 1 strain Mahoney was a gift from H. Ploegh. Venezuelan equine encephalitis virus TC-83 (pVEEV/GFP) was a gift from I. Frolov.

qPCR infectivity assays. Cells were plated in 96-well plates and infected with an MOI of 0.1 of virus, unless otherwise stated. Cells were collected as outlined in Ambion Power SYBR Green Cells-to-Ct kit (Ambion 4402954). Cells were collected 8 h after infection with polio virus, 24 h after infection with Sindbis virus, Venezuelan equine encephalitis virus and human rhinovirus 14, 2 days after infection with DENV-2 16681 and Zika virus, yellow fever virus, and West Nile virus Kunjin strain, 3 days after infection with HCV JFH1 and 5 days after infection with DENV-2 New Guinea. For comparison of DENV serotypes, cells were infected at an MOI of 0.01 with DENV-1 Hawaii 1944, DENV-2 New Guinea C 1944, DENV-3 Philippines/H871856 and DENV-4 H241 Philippines 1956 for 2 days. All samples were normalized to 18S expression. Two independent experiments were performed with triplicate infections and one representative is shown.

The following qPCR primers were used: DENV-2-forward: 5′-GCCCTTCT GTTCACACCAT-3′, reverse: 5′-GGCTCTGCCAATCAGTTCAT-3′; universal-DENV-forward: 5′-GGTTAGAGGAGACCCCTCCC-3′, reverse: 5′-GGTCTCCT CTAACCTCTAGTCC-3′; yellow fever forward: 5′-GAAATGCCTGCCC TTTATGA-3′, reverse: 5′-GCACATGGCAACAGAAGCTA-3′; Kunjin-forward: 5′-GCTTTGCCACCTCTCTTCAC-3′, reverse: 5′-CGGTTGATGGTTTCC ACTCT-3′; ZIKV-forward: 5′-ACCATACGGCCAACAAAGAG-3′, reverse: 5′-TCCACAGCCAGGAAGAGACT-3′; HCV-forward: 5′-TCTCTCAGTCC TTCCTCGGA-3′, reverse: 5′-AAGCCGGCTAGAGTCTTGTT-3′; SINV-forward: 5′-CGCGGTCACGTAAGGATAAT-3′, reverse: 5′-TTTGGCATTCTTCAGC ACAG-3′; polio-forward: 5′-CAACCTCCCACTGGTGACTT-3′, reverse: 5′-ATTTCCCCTGCTCAACCTTT-3′; 18S-forward: 5′-AGAAAC GGCTACCACATCCA-3′, reverse: 5′-CACCAGACTTGCCCTCCA-3′; VEEV-forward: 5′-CAGGACGATCTCATTCTCAC-3′, reverse: 5′-TCATTCA CCTTGTACCGAACG-3′; HRV-14-forward: 5′-AAGCAATTTGGTGGTCC AAG-3′, reverse: 5′-ACACTGGGGTTTGAAGCACT-3′.

Crystal violet staining. For virus infections, wild-type and knockout cell lines were plated out in either 24- or 96-well plates. Cells were infected with DENV-2 (16681), HCV, SINV or polio using a MOI of 1. Huh7-STT3A-KO-STT3B-KO-pLenti-TRE3G-CMV-STT3B cells were cultured in presence or absence of 25 ng ml^{-1} doxycycline. Cells were incubated for 48–120 h then fixed using 4% formaldehyde in PBS. Cell viability at time of fixation was determined by crystal violet staining.

Plaque-forming units assay. Plaque assays were performed on BHK-21 cells. In brief, BHK-21 monolayers were grown to 80% confluency in 24-well plates and incubated for 1 h at 37 °C in 5% CO_2 with serially diluted virus supernatants from wild-type and mutant HAP1 cells infected with DENV, at a MOI of 0.1 for 48 h. The wells were then overlaid with DMEM, 0.8% Aquacide II (EMD Millipore), and 10% FBS, incubated for 7 days, and fixed with 10% formaldehyde. The cells were then stained overnight with crystal violet. The next day the wells were extensively washed with water then dried, and the resulting plaques were counted and plaque-forming units per ml were calculated. Two independent experiments were performed with triplicate infections and one representative is shown.

Focus-forming units assay. Wild-type and knockout Huh7.5.1 cells were plated in 24-well plates and infected with HCV at a MOI of 0.1. Three days after infection, supernatant was collected and added to wild-type Huh7.5.1 cells in a tenfold dilution series. After 3 days, cells were fixed, stained with mouse-anti-HCV-core (Abcam ab2740) and anti-mouse-IgG-Alexa-488 (Life Technologies) and fluorescent colonies were counted. Two independent experiments were performed with triplicate infections and one representative experiment is shown.

Luciferase reporter virus assays. Cells were plated out in 96-well plates in triplicates and infected with dengue luciferase reporter virus at an MOI of 0.01. Cells were incubated with dengue luciferase reporter virus at 37 °C, 5% CO_2 and cell lysates were collected. Luciferase expression was measured using *Renilla* Luciferase Assay system (Promega E2820). Cells were lysed using *Renilla* lysis buffer and luciferase activity measured by addition of substrate and immediate luciferase readings were taken using Glomax 20/20 luminometer using a 10-s integration time. For the cross-comparison of the effects of host factors on DENV and HCV Huh7.5.1 knockout cell lines were infected with dengue luciferase reporter virus at an MOI of 0.01 or with HCV luciferase virus at an MOI of 0.2. For the validation of HCV host factors, four different knockout cell lines (created using lentiCRISPRv2) were infected with HCV luciferase virus at an MOI of 0.2. Two independent experiments were performed with triplicate infections and one representative experiment is shown, with the exception of the experiment shown in Extended Data Fig. 4e, which was performed once with triplicate infections.

Infection of Raji DC-SIGN cells. Raji DC-SIGN host factor knockout cell lines were created by transduction of lentiCRISPRv2 and subsequent puromycin selection. Resulting cell lines were infected with dengue luciferase virus at an MOI of 0.05 and collected 3 days after infection. Three independent experiments were performed with triplicate infections and one representative is shown.

Internalization assay and confocal microscopy. Approximately 10,000 Huh7 cells were seeded on Laboratory-TekII Chamber slides (Thermo Fisher Scientific). The next day, cells were incubated on ice for 15 min, infected with DENV (MOI = 60) and incubated on ice for 1 h. Cells were washed three times with ice-cold PBS and subsequently incubated at 37 °C for 0 or 30 min. At each time point, 10 μg ml^{-1}

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

wheat germ agglutinin-Alexa-594 (Life Technologies, W11262) was added for 10 min at room temperature before three washes with PBS and fixation with 4% paraformaldehyde. Dengue virus was stained with a rabbit-anti-dengue-envelope antibody (Genetex, 127277) in block/perm buffer (1% saponin, 1% Triton X-100, 5% FBS) for 1 h followed by incubation with goat anti-rabbit-IgG-Alexa-488 (Life Technologies, A-11008) and DAPI (Insitus, F203) for 30 min. After three washes with PBS, cells were visualized using confocal microscopy.

Replicon assays. Dengue replicon plasmid was linearized using XbaI restriction enzyme. Replicon RNA was generated using the MEGAscript T7 High Yield Transcription Kit (Ambion, AM1334) with the reaction containing 5mM $m^7G(5')ppp(5')G$ RNA Cap Structure Analogue (NEB, S1405S). Resulting RNA was purified by sodium acetate ethanol precipitation. HCV sgJFH1 replicon[33] RNA was prepared as described for DENV with the exception of adding the cap structure analogue. Cells were washed twice with PBS and re-suspended in electroporation buffer (Teknova, E0399). Three micrograms of purified replicon RNA was mixed with cells, and cells were electroporated using Bio-Rad Gene Pulser Xcell electroporator using square wave protocol. Electroporated cells were resuspended in cell culture medium without antibiotics and plated into 24-well plates. Luciferase expression was measured using Renilla Luciferase Assay system (Promega, E2820). Cells were lysed using *Renilla* lysis buffer and luciferase activity measured by addition of substrate and luciferase readings were taken immediately using Glomax 20/20 luminometer using a 10-s integration time. For lumiflavin treatment, cells were electroporated with 2 μg of viral RNA and 1 μg of firefly mRNA (Trilink) to normalize for effects on cell proliferation. For the lumiflavin treatment, two independent experiments with three electroporations each were performed. One representative experiment is shown. For the replicon assay in *ELAVL1*-knockout cells, three independent experiments with a single electroporation were performed. The average of the experiments is shown. For the DENV-replicon assay, three independent experiments were performed. One representative experiment was shown.

Immunoblot analysis. Cell pellets were lysed using Laemmli SDS sample buffer containing 5% β-mercaptoethanol and boiled for 10 min. Lysates were separated by SDS–PAGE on pre-cast Bio-Rad 4–15% poly-acrylamide gels in Bio-Rad miniprotean gel system. Proteins were transferred onto PVDF membranes using Bio-Rad trans-blot protein transfer system. PVDF membranes were blocked with PBS buffer containing 0.1% Tween-20 and 5% non-fat milk. Blocked membranes were incubated with primary antibody diluted in blocking buffer and incubated overnight at 4 °C rotating. Primary antibodies were detected using horseradish peroxidase (HRP)-conjugated secondary anti-mouse and anti-rabbit antibodies (Genetex) by incubating membranes at a 1:5,000 dilution for 1 h at room temperature. Antibody-bound proteins were detected by incubating with Pierce West Pico and Extended Duration Peroxide Solutions and visualized on film. Wild-type cells were treated with 10 μg ml^{-1} tunicamycin and treated for 3–4 h at 37 °C, 5% CO_2. To visualize proteins by immunoblotting, the following antibodies were used anti-SHBG (Genetex, GTX63795) at a dilution of 1:2,500. Anti-pSAP (Genetex, GTX101064) at a dilution of 1:2,500. Anti-HA C29F4 (Cell Signaling, 3724P) at a dilution of 1:2,500. Anti-mouse M2-Flag (Sigma, F1804-200UG) at a dilution of 1:2,500. Anti-DYKDDDDK (Flag) (Cell Signaling, 2368) at a dilution of 1:2,500. Anti-RPN1 (gift from M. Ivessa) at a dilution of 1:2,000. Anti-NS1 (Genetex GTX124280) at a dilution of 1:2,500. Anti-P84 (Genetex GTX70220) at a dilution of 1:3,500. Anti-DENV-ENV (Genetex GTX127277) at a dilution of 1:2,500. Anti-prM (Genetex, GTX128092) at a dilution of 1:2,500. Anti-NS2B (Genetex, GTX124246) at a dilution of 1:2,500. Anti-NS3 (Genetex, GTX124252) at a dilution of 1:2,500. Anti-RFK (Sigma, SAB1409492) at a 1:500 dilution. Anti-FLAD1 (Santa Cruz Bio, sc-376819) at a 1:250 dilution. Anti-STT3B (Sigma, HPA036646) at a dilution of 1:1,000. Anti-MAGT1 (Proteintech Group, 17430-1-AP) at a dilution of 1:1,000. Anti-RPS25 (Abcam, 102940) at a dilution of 1:1,000. Anti-HUR (Santa Cruz, sc-5261) at a dilution of 1:200. Anti-SRRD (Sigma, HPA002945) at a dilution of 1:500.

Lentiviral or retroviral complementations. Lentiviral or retroviral transduction was used to create stable cell lines expressing a selected gene of interest. Respective genes of interest (see 'Construction of lenti- or retroviral constructs' section) were cloned into the pLenti-CMV-Puro-DEST vector (w118-1) (a gift from E. Campeau), or PMX-IRES-BLAST-DEST. Lentivirus or retrovirus produced in HEK293FT cells was used to transduce respective cell lines overnight. Cells stably expressing the gene of interest were selected by treatment with 1–4 μg ml^{-1} puromycin or 10–50 μg ml^{-1} blasticidin over 2 days (InvivoGen) along with untransduced cells as negative control.

Genome engineering. CRISPR guide RNA sequences were designed using the Zhang laboratory CRISPR design tool (see Extended Data Fig. 3 for CRISPR target sites). Corresponding oligonucleotides or geneblocks containing U6 promoter sequence and U6 termination sequence were ordered from IDT. Oligonucleotides were cloned into the Zhang laboratory generated Cas9 expressing pX458 guide RNA plasmid (Addgene) as previously described using Gibson assembly reaction New England Biolabs. Geneblocks were cloned into pCR-Blunt II-TOPO vector (Life Technologies). TOPO-cloned geneblocks were co-transfected into respective cells with a mCherry-expressing construct and hCas9-expressing vector (Addgene 41815 hCas9 Church pcDNA3.3-Topo) guide RNA encoded in the pX458 plasmids were transfected alone using Lipofectamine 2000 (Life technologies) according to manufactures guidelines. Transfected cells were single cell sorted based on GFP or mCherry expression into 96-well plates using BD influx cell sorter. Clonal cell lines were allowed to expand and genomic DNA was isolated for sequenced based genotyping of targeted allele. For this, a 500–700-base pair (bp) region that encompassed the guide RNA-targeted site was amplified and the PCR product was Sanger sequenced. In haploid cells (HAP1), only one mutated allele was present in the sequenced PCR product and cellular subclones containing a frame shift mutations or large indels were selected. In aneuploid Huh7.5.1 cells, we regularly observed that the PCR product contained more than 1 trace, suggesting non-identical mutations in multiple alleles. In this case, the PCR product was cloned into a plasmid vector and colonies were sequenced to separate allele specific mutations. Subclones were chosen where all alleles were mutated. It should be noted that in aneuploid Huh7.5.1 cells, we sometimes observed cellular subclones where all mutant alleles contained the same mutation (for example, CD81 and ELAVL1). It has been reported that CRISPR/Cas9 technology can generate homozygous bi-allelic mutations more frequently than expected in diploid cells or cancer cells[34,35], perhaps because both alleles were independently repaired in an identical manner or because one allele served as a template for homology-directed repair of the other allele. To create KO cell lines using lentiCRISPRv2 (Addgene) the following guide RNAs below were cloned into the vector, Huh7.5.1 cells were lentivirally transduced and selected with Puromycin. The following guide RNA sequences were used: ANKRD49: AGAAAGGAGTCTCCGCACTG; ANKRD49 guide2: ATGAACCGTTACGTCAAACC; ANKRD49 guide3: GCCCAAAGAAGCAATCTGCT; ANKRD49 guide4: AGAAAGGAGTCTCCGCACTG; CD81: GCGCCCAACACCTTCTATGT; CLDN1: CGATGGCGCCGATCCATCCC; ELAVL1: TTGGGCGGATCATCAACTCG; ELAVL1 guide2: TGTGAACTACGTGACCGCGA; ELAVL1 guide3: GGGCCTCCGAACCGTCGCGC; ELAVL1 guide4: AGAGCGATCAACACGCTGAA; EMC1: AGGCCGAATCATGCGTTCCT; EMC2: GATTGCCATTCGAAAAGCCC; EMC3: GTGCCACCTTCTCCTATGAC; EMC4: TGCTTGTCCAAGTAACCGAC; FKBPL: GTCAAGAAGATCGTAATCCG; FKBPL guide2: GAAGAGCCCGTCCATAGCAT; FKBPL guide3: ACAGAGCTAACTATGGGCGT; FKBPL guide4: GTTTCGGTAGGAGGGTCTCG; FLAD1: ACAGACCATTGAGACCTCCC; FLAD1 guide2: CATGCGCATCAACCCACTGC; FLAD1 guide3: TACAGGAGTAGGGGTCAGTC; FLAD1 guide4: TGTGTCCCTGGGGGTTGAAG; MAGT1: GAGCGAACATGGCAGCGCGT; MIR-122: GAGTTTCCTTAGCAGAGCTG; MMGT1: CAGGCACTTACGCTGCGCAG; non-targeting: GCCCAGACGCCCTAGAATAG; OCLN: ACGTAGAGTCCAGTAGCTGC; OSTC: TCAGTCATAGAACCGACACT; PPIA: GTACCCTTACCACTCAGTCT; RFK: TATCATGCATACCTTCAAAG; RFK guide2: GGTCAAGTGGTGCGGGGCTT; RFK guide3: CTATGGGGAAATCCTCAATG; RFK guide4: CCAACCATAGTAAATACCAG; RPN2: TCGCTACCACGTGCCAGTTG; SRRD: GACTGTTCTCAGTGAGAACG; SRRD guide2: GATAGATACCTTTGCAATGT; SRRD guide3: ATTGAAGTCCTTAACACCCT; SRRD guide4: AACAACTGAAGGCCCCTGTG; SSR2: CAATAGCAGGGGGATGCCGA; SSR3: GACCCTAGTAAGCACATATT; STT3A: GTACTCACGGATCAAACTCA; STT3B: TACAGCAAAAGAGTCTACAT; ZEB1: TGAAGACAAACTGCATATTG.

TALENs targeting *AUP1* in HAP1 cells were generated as indicated in Extended Data Fig. 3. Cells were co-transfected with left and right TALEN-containing constructs and an mCherry-expressing construct using Lipofectamine 2000 (Life Technologies) according to manufacturer's guidelines. Transfected cells were single-cell sorted based on mCherry expression into 96-well plates using BD influx cell sorter. Subclones were allowed to expand and genomic DNA was isolated for sequenced based genotyping of *AUP1* allele. HAP1 cells containing gene-trap insertions in *STT3A*, *STT3B*, *RPN1*, *SSR2*, *SSR3*, *ASCC2* and *RPS25* were isolated by picking resistant colonies from the DENV-2 haploid genetic screen. Picked colonies were screened for gene-trap insertions using PCR with primers directed to the gene-trap and the flanking region of the gene of interest.

Co-immunoprecipitation. Wild-type HAP1 cells were transduced with STT3A–Flag, STT3B–Flag or RPS25–Flag lentiviral vectors (see 'Construction of lenti- or retroviral constructs' section). HAP1 cells expressing Flag-tagged proteins were trypsinized and washed once with PBS. Cells were lysed with TNM buffer (25 mM Tris-HCl, 15 mM NaCl, 5 mM $MgCl_2$) containing 1% digitonin, 1 mM PMSF, and Halt Protease and Phosphatase Inhibitor Cocktail (Life Technologies, 78440) for 1 h on ice gently vortexing every 15 min. Cell lysates were clarified by centrifugation at 15,000*g* for 10 min. Clarified lysates were incubated at 4 °C overnight

UNCORRECTED PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

with Anti-FlagG M2 Magnetic Beads (Sigma M8823-5ml). Incubated beads were washed three times with TNM buffer containing 0.1% digitonin, 1 mM PMSF, 1× halt protease and phosphatase inhibitor. Cells were washed once with TNM buffer and competitively eluted with TNM buffer containing 150 ng μl^{-1} 3× Flag Peptide (Sigma, F4799) for 30 min on ice. For immunoblotting elutions were denatured by boiling in 5× sample buffer and analysed by SDS–PAGE using antibodies against DENV non-structural proteins. Elutions were also prepared for mass spectrometry analysis. Cells expressing RPS25–Flag, a host protein with a likely different mechanistic action, as well as untransduced cells, were used as a negative control in these experiments.

SYPRO ruby staining. After electrophoresis gel was fixed for 30 min in fixative buffer (50% methanol, 7% acetic acid) and incubated with SYPRO Ruby Stain (Fischer, S-12000) overnight. Gels were then washed once with wash buffer (10% methanol, 7% acetic acid) and twice with distilled water. Gel was imaged on Molecular Dynamics Storm scanner.

Construction of STT3A/STT3B double-knockout cell line with STT3B conditionally expressed using a Tet-On system. HAP1 cells stably transduced with the transactivator pLenti-CMVrtTA3G-Blast (R980-M38-658) (Addgene, 31797) with the endogenous *STT3B* gene disrupted by CRISPR/Cas9 were lentivirally transduced with pLenti-CMVTRE3G-Puro-STT3B-Flag, which drives STT3B under a doxycycline-inducible promoter. Transduced cells were then transfected with pX458 plasmid encoding a guide RNA targeted to the *STT3A* gene. Transfected cells were then subcloned based on GFP expression of PX458 plasmid into 96-well plates containing IMDM plus 10% FBS, penicillin-streptomycin, L-glutamine and 25 ng ml^{-1} doxycycline. Subclones were allowed to grow for 2 weeks, then replica plated and in one replicate the doxycycline medium was washed away and replaced with regular growth medium without doxycycline and incubated for 5 days. Cells that were dependent on doxycycline for growth were genotyped to verify both endogenous *STT3A* and *STT3B* had double frame shifting CRISPR/Cas9 editing events. *STT3A*/*STT3B* endogenous double knock out cells were then lentivirally transduced with wild-type or mutant STT3A or STT3B under the CMV promoter. The lentivirally transduced cell lines were then plated in 96-well plates and the doxycycline was washed away and incubated for 5 days. Cells were then fixed and stained with crystal violet to assess cell viability.

Treatment of HCV infected cells with lumiflavin, FMN and FAD. Wild-type Huh7.5.1 cells were treated with lumiflavin (Santa Cruz Bio, sc-224045) ranging from 10 to 100 μM and infected with HCV or DENV-2 at a MOI of 0.1. For WNV, YFV, PV (polio virus), SINV, VEEV and HRV-14 cells were treated with 50 μM lumiflavin and infected at a MOI of 0.1. For rescue of HCV replication in lumiflavin-treated cells 100 μM FMN and 10 mM FAD were used. RFK- and FLAD1-KO Huh7.5.1 were cultured in absence or presence of 500 μM FMN (TCI America, R0023) or FAD (Sigma, F8384) and subsequently infected with HCV at a MOI of 0.1. After 3 days of infection, levels of infection were determined using immunofluorescence, western blot and qPCR. Anti-HCV core 1b (Abcam ab2740) was used at 1:500 for immunofluorescence and 1:1,000 for western blotting. Anti-DENV-2 NS5 (GeneTex GTX103350) was used at a 1:25,00 dilution for western blotting. For the lumiflavin treatment two independent experiments with triplicate infections were performed and one representative is shown. For the FMN/FAD complementation two independent experiments were performed and the average is shown.

MTT assay. To test effects of lumiflavin on cell viability MTT assay was performed according to the manufacturer's instructions (Sigma Cell Proliferation Kit I (MTT), 11465007001). Three independent experiments in triplicates each were performed and one representative is shown.

Transmission electron microscopy. Cells stably expressing STT3B-APEX2 were plated in 6-well plates and infected with an MOI 5 of DENV-2. 28 h after infection, cells were washed with PBS and fixed with 2% glutaraldehyde in 100 mM sodium cacodylate, 2 mM $CaCl_2$, pH 7.4, buffer. Cells were fixed at 4 °C for 60 min then washed three times with PBS. Fixed cells were quenched with 100 mM sodium cacodylate, 2 mM $CaCl_2$, pH 7.4, and 20 mM glycine. Quenched cells were washed twice with PBS and stained with using the KPL DAB reagent set (KPL, 54-10-00) for 8 h. After incubation with DAB, cells were rinsed twice with PBS and scraped off well using a cell scraper and pelleted. Pelleted cells were re-suspended in 10% gelatin in 0.1 M sodium cacodylate buffer, pH 7.4, at 37 °C and allowed to equilibrate for 5 min. Cells were pelleted again, excess gelatin removed, then chilled in cold blocks and covered with cold 1% osmium tetroxide (EMS, 19100) for 2 h rotating in a cold room. They were then washed three times with cold ultrafiltered water, then en bloc stained overnight in 1% uranyl acetate at 4 °C while rotating. Samples were then dehydrated in a series of ethanol washes for 20 min each at 4 °C beginning at 30%, 50%, 70%, 95% where the samples were then allowed to rise to room temperature, changed to 100% ethanol twice, then propylene oxide for 15 min. Samples were then infiltrated with EMbed-812 resin (EMS, 14120) mixed 1:2, 1:1 and 2:1 with propylene oxide for 2 h each with leaving samples in 2:1 resin to propylene oxide overnight rotating at room temperature in the hood. The samples were then placed into EMbed-812 for 2–4 h then placed into molds with labels and fresh resin, orientated and placed into a 65 °C oven overnight. Sections were taken at approximately 80 nm, picked up on formvar/carbon-coated 100-mesh Cu grids, stained for 30 s in 3.5% uranyl acetate in 50% acetone followed by staining in 0.2% lead citrate for 3 min. Observed in the JEOL JEM-1400 120kV and photos were taken using a Gatan Orius 4k X 4k digital camera.

Mass spectrometry

Liquid chromatography-tandem mass spectrometry. Elutions from co-immunoprecipitations were trypsin digested and purified using Sep Pak C18 purification column. Peptides were analysed using an LTQ Velos Orbitrap mass spectrometer (Thermo Fisher Scientific) coupled to an Agilent 1100 high performance liquid chromatography pump (Agilent Technologies) and a MicroAS autosampler (Thermo Fisher Scientific). Peptide mixtures were introduced into the mass spectrometer via a fused silica microcapillary column (100 μm inner diameter) ending in an in-house pulled needle tip (internal diameter ≈ 5 μm). Columns were packed to a length of 17 cm with a C18 reversed-phase resin (Magic C18AQ; Michrom Bioresources). Peptides were loaded onto the column and then eluted into the nanospray ionization source of the mass spectrometer via a two-step gradient of 7–25% buffer B (2.5% water and 0.1% formic acid in acetonitrile (v/v)) in buffer A (2.5% acetonitrile and 0.1% formic acid in water (v/v)) over 60 min followed by a second phase of 25–45% buffer B over 20 min. Eluting peptides were measured by the LTQ Velos Orbitrap operating in a data-dependent mode in which 10 ion-trap MS/MS spectra were acquired per data-dependent cycle from a high-resolution ($R = 60{,}000$) precursor spectrum (mass range = 360–1,600 m/z).

Mass spectrometry data processing. Raw data files produced by the mass spectrometer were converted to the mzXML format using in house software, MS Convert. MS and MS/MS data were extracted from mzXML files with in-house software. MS/MS spectra were analysed using Sequest algorithm searching a composite target-decoy protein sequence database. The target sequences consisted of human proteins downloaded from the Uniprot database (11-17-2014) and protein sequence corresponding to the dengue virus 2 16681 polyprotein. Decoy sequences were created by reversing the orientation of all target sequences. Parameters used for all searches included the requirement of trypsin peptide cleavage, two missed cleavages allowed, peptide mass tolerance of 20 p.p.m., variable oxidation of methionine residues (+15.99491 Da), and static carbamylation modification of cysteine residues (+57.02146). Decoy peptide identifications guided the creation of filtering criteria delivering preliminary sets of peptide-spectrum matches with estimated false discovery rate <1%. Spectral counts for each condition were combined at a protein level and normalized by protein length to infer protein abundances in each case.

DENV reporter virus and DENV replicon design and generation

Construction of pDENV-Luc replicon. The design of the DENV replicon was based on DVRep described previously[36]: The viral 5′ untranslated region (UTR) was followed by a duplication of the first 102 nucleotides of the C coding region, which contain cis-acting elements required for replication (CAE). The CAE was fused to the renilla luciferase coding region followed by the DENV open reading frame (ORF) starting at the signal peptide preceding NS1. Between the luciferase and the DENV structural proteins a foot and mouth disease virus (FMVD) 2A sequence was introduced to provide cotranslational cleavage and release of luciferase. The construct was based on pD2/IC-30P, which contains a full-length infectious clone encoding DENV-2 strain 16681 (ref. 37). We also included the amino acid mutation Q399H in the envelope protein. We gene-synthesized a fragment containing the T7 polymerase promoter sequence followed by the first 102 nucleotides of the C coding region in frame with *Renilla* luciferase and FMDV 2A followed by the DENV open reading frame (ORF) starting at the signal peptide preceding NS1 until an internal HpaI site. This fragment was released by SacI (preceding the T7 promoter) and HpaI and cloned in pD2/IC-30P in a three point ligation with KpnI/SacI and KpnI/HpaI fragments.

Construction of pDENV-Luc infectious clone. The design of the DENV reporter was based on mDV-R described previously[38]: The viral 5′ UTR was followed by a duplication of the first 104 nucleotides of the C coding region, which contain cis-acting elements required for replication (CAE). The CAE was fused to the renilla luciferase coding region followed by the complete DENV ORF. Between the luciferase and the DENV structural proteins a FMVD 2A sequence was introduced to provide cotranslational cleavage and release of luciferase. The construct was based on pD2/IC-30P, which contains a full-length infectious clone encoding dengue virus serotype 2 strain 16681 (ref. 37) in which an envelope Q399H mutation was introduced that enhanced viral infection in mammalian cells using primers 5′-GGAAGTTCTATCGGCCACATGTTTGAGACAAC-3′ and 5′-GTTGTCTCAAACATGTGGCCGATAGAACTTCC-3′ via the QuikChange Site-Directed Mutagenesis kit (Agilent Technologies). We gene-synthesized a fragment containing the T7 polymerase promoter sequence

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

followed by the first 104 nucleotides of the C coding region in frame with *Renilla* luciferase and FMDV 2A. This fragment was PCR amplified, introducing a SacI site at the 5′ end and a NheI site (present in the FMDV 2A sequence) at the 3′ end using primers: 5′-CGAAATTCGAGCTCACGCG-3′ and 5′-TCCTGCTAGCTTGAGCAAATCAAAGTTC-3′. To create and in frame fusion of FMDV 2A with the DENV ORF a second DNA fragment was amplified using pD2/IC-30P as template with primers: 5′-TCAAGCTAGCAGGAGACGT TGAGTCCAACCCCGGGCCCATGAATAACCAACGGAAAAAGGCG-3′ and 5′-GGAAGAGCATGCAGTCGGAAATG-3′ introducing 5′ NheI and 3′ SphI restriction sites. The two fragments were cut with the respective restriction enzymes and ligated into pD2/IC-30P cut with SacI and SphI to create pDENV-Luc. DENV-Luc virus was produced by cutting with XbaI to linearize plasmid and *in vitro* transcription performed of pDENV-Luc and transfection into BHK cells using Lipofectamine 2000.

Construction of lenti- or retroviral constructs. PMX-IRES-BLAST-DEST was made by cutting pMXs-IRES-Blasticidin Retroviral Vector (Cell Biolabs, RTV-016) with SnaBI and the Gateway destination cassette (reading frame A) was blunt cloned in to this vector according to manufacturer's protocol (Gateway Vector Conversion System; Invitrogen, 11828-029)

To generate a lentiviral construct expressing STT3A–Flag, Dharmacon cDNA BC020965 was used as template to generate a PCR product using primers 5′-CACCATGACTAAGTTTGGATTTTGCG-3′ and 5′-TTACTT ATCGTCGTCATCCTTGTAATCTGTCCTTGACAAGCCTCGATT-3′

Amplified PCR product was then topo cloned into Gateway compatible entry vector pENTR/D-TOPO Cloning Kit (Life Technologies K2400-20) and gateway reaction (Life Technologies) used to insert into pLenti-CMV-Puro-Dest (w118-1).

To generate a lentiviral construct expressing STT3B–Flag, Dharmacon cDNA BC052433 was used as template to generate PCR product using primers forward primer 5′-CACCATGTCTTGGTGGGATTATGGC-3′ and reverse primer 5′-TTA CTTATCGTCGTCATCCTTGTAATCAACAGTCTTCTTAGAGGTCTTCTT-3′. It should be noted that we used *Mus musculus* STT3B because we were unable to clone human STT3B.

Amplified PCR product was then TOPO cloned into Gateway compatible entry vector pENTR/D-TOPO Cloning Kit (Life Technologies, K2400-20) and gateway reaction was used to insert into pLenti-CMV-Puro-Dest (w118-1).

To generate an SHBG-expressing construct, SHBG was ordered as two geneblocks and used to generate PCR product with primers 5′-TGTG GTGGAATTCTGCAGATACCTGTGGTGGAATTCTGCAGATACC-3′ and 5′-ATCCAGCACAGTGGCGG-3′. PCR product was Gibson cloned Gibson assembly reaction kit (New England Biolabs) into pLenti-CMV-Puro-Dest (w118-1) that was EcoRV digested.

To generate a Flag3×–RPS25 expression construct, entry vector PENTR-Flag3×–RPS25 was generated as described[39] using the forward primer CACCATGGACTACAAAGACCATGACGG. PENTR-Flag3×–RPS25 was then used in a Gateway reaction (Life Technologies) to introduce Flag3×–RPS25 into PMX-IRES-BLAST-DEST retroviral expression construct.

A construct expressing STT3B fused with APEX2 and Flag-tagged PCR products were generated using the pLenti-STT3B expression construct described above and APEX2 Addgene plasmid 49386 (refs 40, 41) as templates to generate PCR products using primers 5′-GACTCTAGTCCAGTGTGGTG-3′ with 5′-AACAGTCTTCTTAGAGGTCTTC-3′ and 5′-GAAGACCTCTAAG AAGACTGTTATGGACTACAAGGATGACGA-3′ with 5′-CGGCCGCCACT GTGCTGGATTTAGGCATCAGCAAACCCAAG-3′. PCR products were Gibson assembled Gibson assembly reaction kit (New England Biolabs) into pLenti-CMV-Puro-Dest (w118-1) that was EcoRV digested.

To generate a STT3B-doxycycline-lenti construct, Dharmacon cDNA BC052433 was used as a template to generate PCR product using primers 5′-CACCATGTCTTGGTGGGATTATGGC-3′ and 5′-TTACTTATCGTCGTCA TCCTTGTAATCAACAGTCTTCTTAGAGGTCTTCTT-3′

Amplified PCR product was then TOPO cloned into Gateway compatible entry vector pENTR/D-TOPO Cloning Kit (Life Technologies, K2400-20), and gateway reaction was used to insert into the doxycycline inducible lentiviral vector pLenti CMVTRE3G Puro DEST (w811-1) (Addgene, 27565).

To generate catalytic site mutants (Extended Data Fig. 5 and refs 15, 42), STT3A- and STT3B-expressing constructs, DNA fragments were generated using pLenti-STT3B-Flag described above as a template using pLenti-EcoRV primers and mutant primers to generate two PCR products (see primers below). Both PCR products for each mutation were Gibson cloned (Gibson assembly reaction kit; New England Biolabs) into pLenti-CMV-Puro-Dest (w118-1) that was EcoRV digested.

Primers forward (F) and reverse (R) were as follows: pLenti EcoRV-F 5′-GACTCTAGTCCAGTGTGGTG-3′, pLenti EcoRV-R 5′-ATCCAGAGGTTGAT TGTCGAG-3′. STT3A mutations: E63A-F 5′-CAGGTTCCTGGCTGAGGC CGGGTTTTATAAATTCCATAACTGG-3′, E63A-R 5′-CCGGCCTCAGCC AGGAACCTG-3′; D167A-F 5′-CTGTGGCTGGCTCCTATGCCAATGA AGGGATTGCCATCTTTTG-3′, D167A-R 5′-CATTGGCATAGGAGCCAGCC ACAG-3′; E351Q-F 5′-CCATCATTGCTTCTGTGTCTCAGCATCAGCCC ACAACCTG-3′, E351Q-R 5′-ATGCTGAGACACAGAAGCAATGATGG-3′. STT3B mutations: D100A-F 5′-ATCATCCACGAGTTCGCCCCGTGGTTTAAC TATAG-3′, D100A-R 5′-CTATAGTTAAACCACGGGGCGAACTCGTG GATGAT-3′; D218A-F 5′-CAGTGGCGGGATCCTTTGCCAATGAAGGCATTG CCATT-3′, D218A-R 5′-AATGGCAATGCCTTCATTGGCAAAGGATCCCG CCACTG-3′; E402Q-F 5′-CAATTATTGCATCAGTGTCTCAGCATCAGCCTA CGACATGG-3′, E402Q-R 5′-CCATGTCGTAGGCTGATGCTGAGACAC TGATGCAATAATTG-3′.

ELAVL1 fused with Flag cDNA was prepared from total RNA of Huh7.5.1 cells using Biorad RT Superscript, and ELAVL1 was PCR amplified with 5′-TGTGGTGGAATTCTGCAGATACCATGTCTAATGGTTATGAAGACCA-3′ and 5′-CGGCCGCCACTGTGCTGGATTTACTTATCGTCGTCATCCTTGT AATCTTTGTGGGACTTG-3′. Next, using Gibson Assembly the PCR product was cloned into pLenti-CMV-Puro-Dest (w118-1) that was digested with EcoRV.

To generate RPN1 fused to 2Strep cDNA, BC010839 was PCR amplified with 5′-CACCATGGAGGCGCCAGCCGC-3′ and 5′-CTACAGGGCATCCAG GATG-3′. The amplified fragments were cloned into P-ENTR-D-Topo (Invitrogen) then a Gateway LR reaction was performed to shuttle cDNA into pLenti-CMV-puro expression vector.

To generate SSR2 fused to 2Strep cDNA, NM_003145.3 was PCR amplified with 5′-CACCATGAGGCTGCTGTCATTTGTG-3′ and 5′-TCAGTTCTTC TTCGTTTTGGGAG-3′. The amplified fragments were cloned into P-ENTR-D-Topo (Invitrogen) then a Gateway LR reaction was performed to shuttle cDNA into pLenti-CMV-puro expression vector.

To generate SSR3 fused to 2Strep cDNA, NM_003145.3 was PCR amplified with 5′-CACCATGGCTCCTAAAGGCAGCTC-3′ and 5′-CTATTTGGA GCCAGTAGACAG-3′. The amplified fragments were cloned into P-ENTR-D-Topo (Invitrogen) then a Gateway LR reaction was performed to shuttle cDNA into pLenti-CMV-puro expression vector.

To generate ASCC2 fused to 2Strep, BC025368 was PCR amplified with 5′-CACCATGCCAGCTCTGCCCCTGG-3′ and 5′-TCAGGATGGG ATCATGCCTTTGCT-3′. The amplified fragments were cloned into P-ENTR-D-Topo (Invitrogen) then a Gateway LR reaction was performed to shuttle cDNA into pLenti-CMV-puro expression vector.

To generate RPS25 fused to 2Strep, NM_001028 was PCR amplified with 5′-CACCATGGACTACAAAGACCATGACG-3′ and 5′-TTAATTA ACCTCGAGTTTAAACGCG-3′. The amplified fragments were cloned into P-ENTR-D-Topo (Invitrogen) then a Gateway LR reaction was performed to shuttle cDNA into pLenti-CMV-puro expression vector.

To generate the UBE2J1 expression contruct, cDNA provided by R. Kopito was PCR amplified with 5′-TGTGGTGGAATTCTGCAGATACCATG GAGACCCGCTACAACCTG-3′ and 5′-CGGCCGCCACTGTGCTGGATTT ATAACTCAAAGTCAAATATGTATTC-3′. The amplified fragments were cloned into pLenti-CMV-puro expression vector by a Gibson Assembly reaction.

To generate the SEL1L expression construct, cDNA provided by R. Kopito was PCR amplified with 5′-TGTGGTGGAATTCTGCAGATACCATGCGGG TCCGGATAGGGCTG-3′ and 5′-CGGCCGCCACTGTGCTGGATTTAAA GTCTACTTACCAAAACCATG-3′. The amplified fragments were cloned into pLenti-CMV-puro expression vector by a Gibson Assembly reaction.

To generate AUP1, glycerol stocks containing AUP1 cDNA in pENTR entry vector were ordered from Darmacon (OHS5894-99868092), and a Gateway LR reaction was performed to shuttle cDNA into pLenti-CMV-puro expression vector.

Comparison of knockout screens to siRNA screens. To compare the top 30 host factor genes of the knockout screens to results from previous short interfering RNA (siRNA) screens, we used the following data: (1) Sessions *et al.*[27] (supplementary table 2). To rank the list by strength of phenotype, we used the *P* value. If multiple siRNA sequences per gene were present, we used the one with the stronger effect. (2) Krishnan *et al.*[11] (supplementary table 1). To rank the identified DENV host factor, column AM was filtered for 'required by both WNV and dengue'. The remaining genes were sorted by 'pooled siRNA Fold reduction of DENV' (column AN). (3) Tai *et al.*[29] (table S2). To sort by phenotype, we chose the validated genes scoring with at least two siRNAs and ranked by *P* value. (4) Li *et al.*[28] (from dataset S1). To rank the genes, we took the mean of the average normalized percentage infected cells part one or two of the four siRNAs. The top 10 genes based on phenotype as explained above are shown in Extended Data Fig. 10.

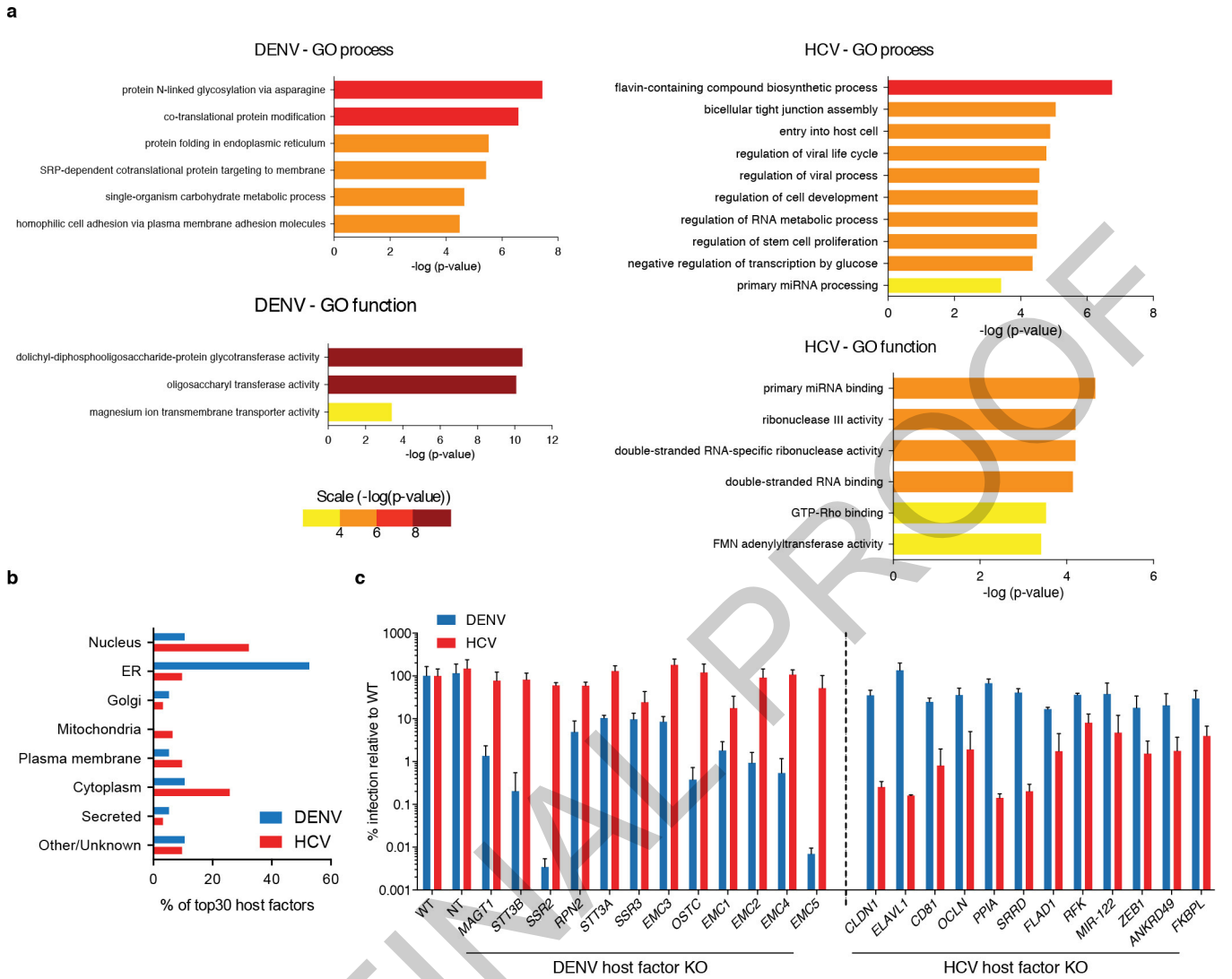
UNCORRECTED FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

31. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11,** 783–784 (2014).
32. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105,** 20380–20385 (2008).
33. Berger, K. L. *et al.* Roles for endocytic trafficking and phosphatidylinositol 4-kinase III alpha in hepatitis C virus replication. *Proc. Natl Acad. Sci. USA* **106,** 7577–7582 (2009).
34. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289,** 21312–21324 (2014).
35. Horii, T. *et al.* Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Sci. Rep.* **4,** 4513 (2014).
36. Alvarez, D. E., De Lella Ezcurra, A. L., Fucito, S. & Gamarnik, A. V. Role of RNA structures present at the 3'UTR of dengue virus on translation, RNA synthesis, and viral replication. *Virology* **339,** 200–212 (2005).
37. Kinney, R. M. *et al.* Construction of infectious cDNA clones for dengue 2 virus: strain 16681 and its attenuated vaccine derivative, strain PDK-53. *Virology* **230,** 300–308 (1997).
38. Samsa, M. M. *et al.* Dengue virus capsid protein usurps lipid droplets for viral particle formation. *PLoS Pathog.* **5,** e1000632 (2009).
39. Fuchs, G. *et al.* Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proc. Natl Acad. Sci. USA* **112,** 319–325 (2015).
40. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12,** 51–54 (2015).
41. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30,** 1143–1148 (2012).
42. Jaffee, M. B. & Imperiali, B. Exploiting topological constraints to reveal buried sequence motifs in the membrane-bound *N*-linked oligosaccharyl transferases. *Biochemistry* **50,** 7557–7567 (2011).

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 1 | Divergence of DENV and HCV host factors. **a**, Gene Ontology (GO) analysis for DENV and HCV CRISPR screens on the ranked gene lists. Curated (by redundancy) enriched GO terms are shown. A complete list of all enriched GO terms can be found in Supplementary Table 4. **b**, Distribution of the subcellular location of the 30 most enriched host factors for DENV and HCV. **c**, Cross-comparison of the effects of DENV or HCV host factor knockout in Huh7.5.1 cells on the replication of DENV or HCV using reporter viruses expressing luciferase. Data are mean and s.d. for triplicate infections.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

a DENV

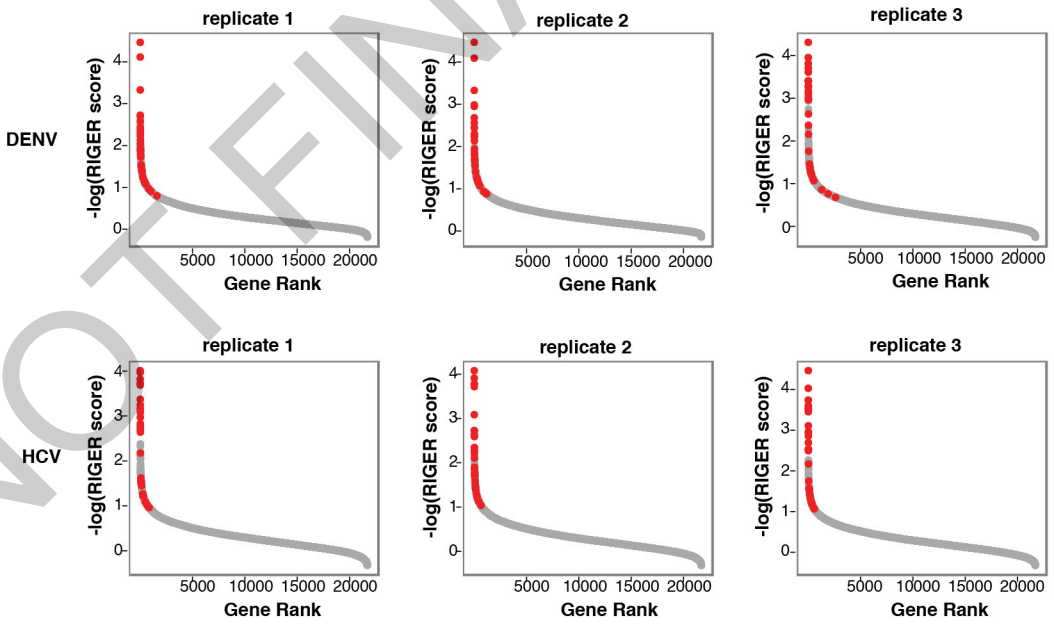
Ranked genes	Rank replicate 1	Rank replicate 2	Rank replicate 3
MAGT1	1	1	1
STT3B	2	2	2
SSR2	3	6	5
RPN2	4	3	13
STT3A	7	4	14
SSR3	5	7	6
EMC3	14	5	12
OSTC	13	9	15
SSR1	11	11	11
EMC1	26	13	9
EMC2	32	14	10
EMC4	10	26	3
EMC5	18	18	8
EMC6	34	24	4
ASCC3	45	131	19
HSPA13	101	74	7
OST4	171	22	22
SEC61A1	88	45	112
DDOST	154	15	168
EMC7	813	290	20
DAG1	1023	237	45
TTC37	413	524	222
CILP	251	46	1286
LYSMD3	47	1061	542
LEPROT	16	42	2615
TTC7A	132	1186	465
ASCC2	399	928	308
TSEN15	511	915	259
SAMD8	147	294	1899
NRSN1	1592	385	220

HCV

Ranked genes	Rank replicate 1	Rank replicate 2	Rank replicate 3
CLDN1	1	2	1
ELAVL1	2	5	2
CD81	4	4	3
OCLN	5	1	4
PPIA	3	3	5
SRRD	6	6	9
FLAD1	8	8	7
RFK	7	7	8
MIR-122	13	9	6
ZEB1	10	13	10
ANKRD49	9	12	15
SGMS1	15	45	11
TRIM26	14	32	89
TECR	98	51	12
CHD7	273	15	13
FKBPL	11	11	516
SND1	292	50	22
DGCR8	19	131	265
BAMBI	87	218	103
TUSC3	246	30	341
MIR-1265	527	28	150
DNAJC22	490	142	80
TMEM248	654	117	48
MSRB2	256	17	578
CBFB	282	78	342
TOR3A	494	419	14
LYZL2	128	287	346
CALML3	61	644	192
CHD1	80	443	280
CTBP1	850	93	85

Top 0.1%
Top 1%
Top 5%

b



Extended Data Figure 2 | Reproducibility of CRISPR screens. a, Ranked lists of the 30 most enriched DENV and HCV host factors and their rankings in the individual replicate screens. The colour code reflects in what percentile the gene scored in the replicate. **b**, Gene enrichment based on RIGER score for the individual replicate screens. Red dots highlight where the 30 most significant host factors ranked in the individual replicates.

NOT FINAL PROOF

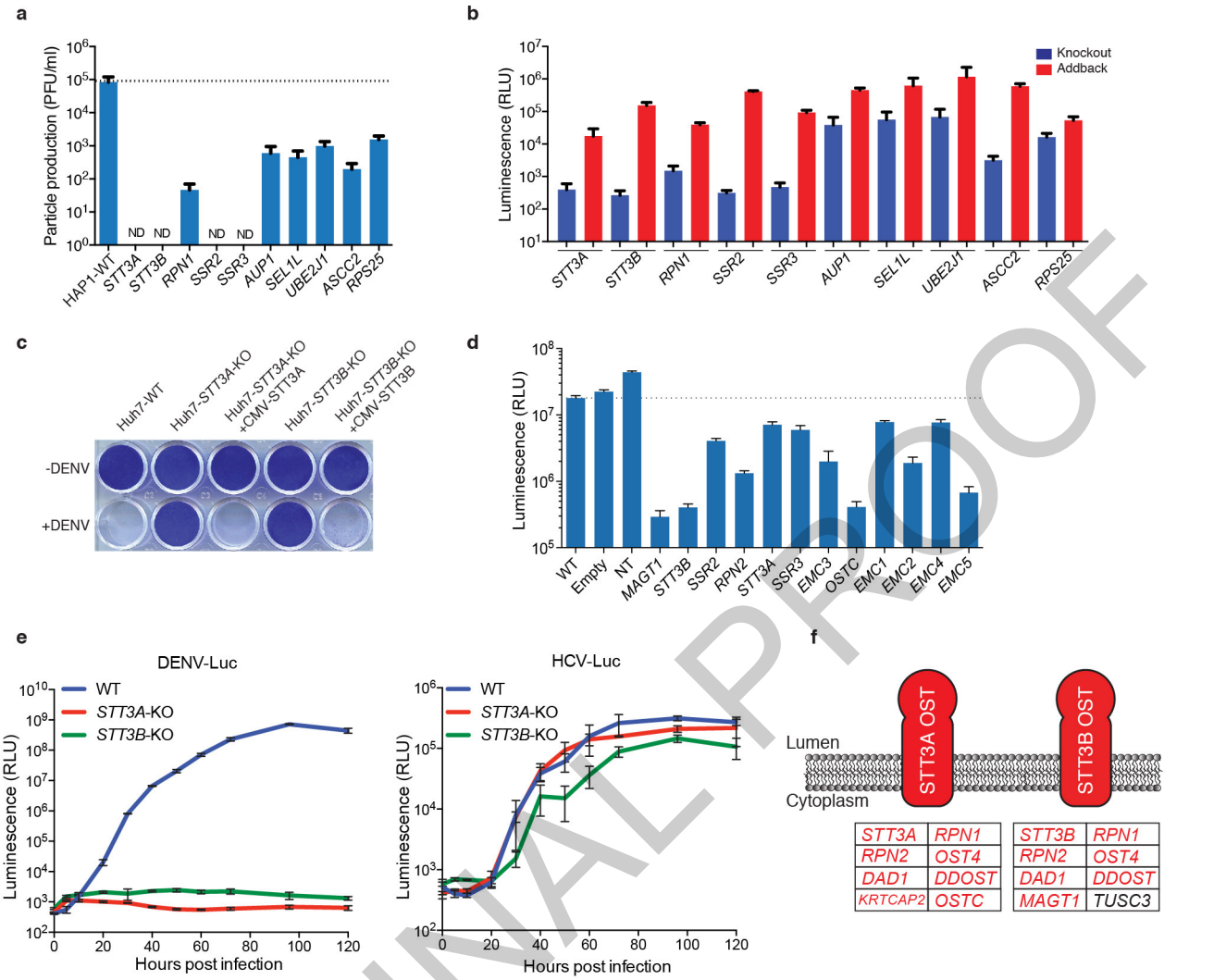
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 3 | Genotyping of cell lines for DENV host factors. a, The site of gene-trap insertion in HAP1 cell lines was determined using a PCR-based method. Bases depicted in red are the flanking sequences upstream of the gene-trap insertion. Bases depicted in green are downstream gene-trap flanking sequences. **b**, TALENs were used to edit the genomic region of *AUP1* in HAP1 cells. Bases depicted in red are the left TALEN-binding site, bases depicted in blue are the right TALEN-binding site. Bases depicted in green are the TALEN target site. Arrow indicates site of 260-bp insertion. **c**, CRISPR Cas9 nuclease was targeted to bases depicted in red in HAP1 cells. Editing events are depicted at the guide RNA target sites below the wild-type sequence. **d**, CRISPR Cas9 nuclease was targeted to bases depicted in red in Huh7 cells. Editing events are depicted at the guide RNA target sites below the wild-type sequence. **e**, Immunoblots of wild type and knockout cell lines.

NOT FINAL PROOF

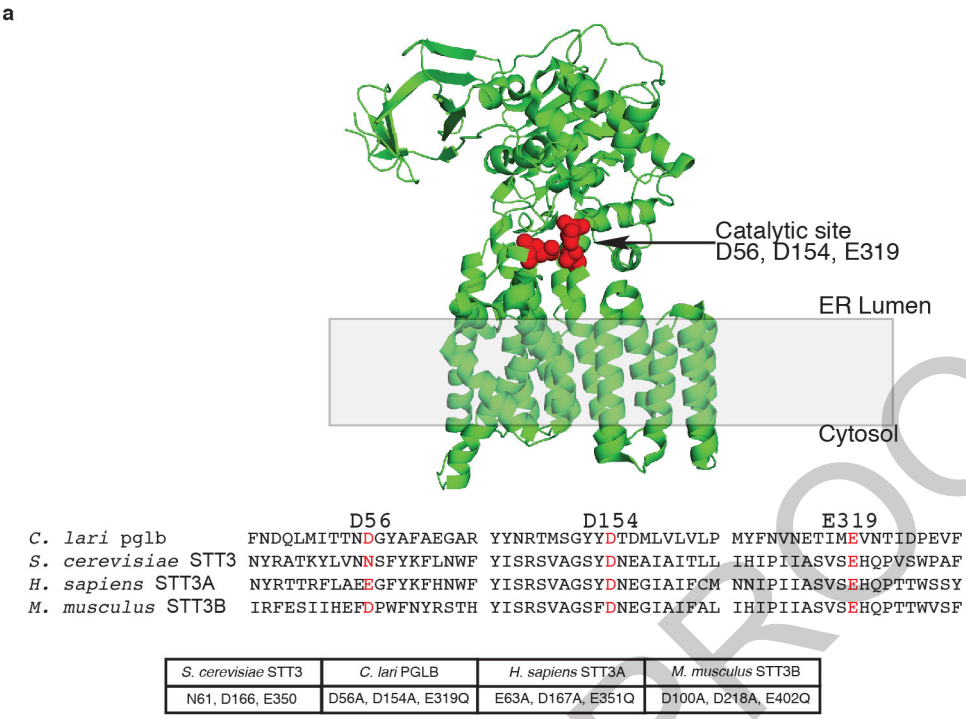
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Validation of DENV host factor genes. **a**, Plaque-forming units (PFU) assay of DENV infection. ND, no plaques detected (threshold of detection is 6 PFU ml^{-1}). **b**, DENV luciferase levels in HAP1 isogenic knockout cells complemented using lentiviral stable expression of corresponding genes. **c**, Crystal violet of complemented Huh7 knockout cells infected with DENV. **d**, DENV luciferase levels in Raji DC-SIGN cells with knockout in DENV host factors (lentiCRISPRv2). Empty denotes an empty vector control (expressing Cas9 but no guide RNA), and NT denotes a cell line expressing a non-targeting guide RNA. **e**, Time course of DENV and HCV expressing *Renilla* luciferase in Huh7 knockout cells. **f**, Schematic diagram of the STT3A and STT3B isoforms. Gene names in red indicate OST subunits identified in the DENV screens. Data are mean and s.d. for triplicate infections.

NOT FINAL PROOF

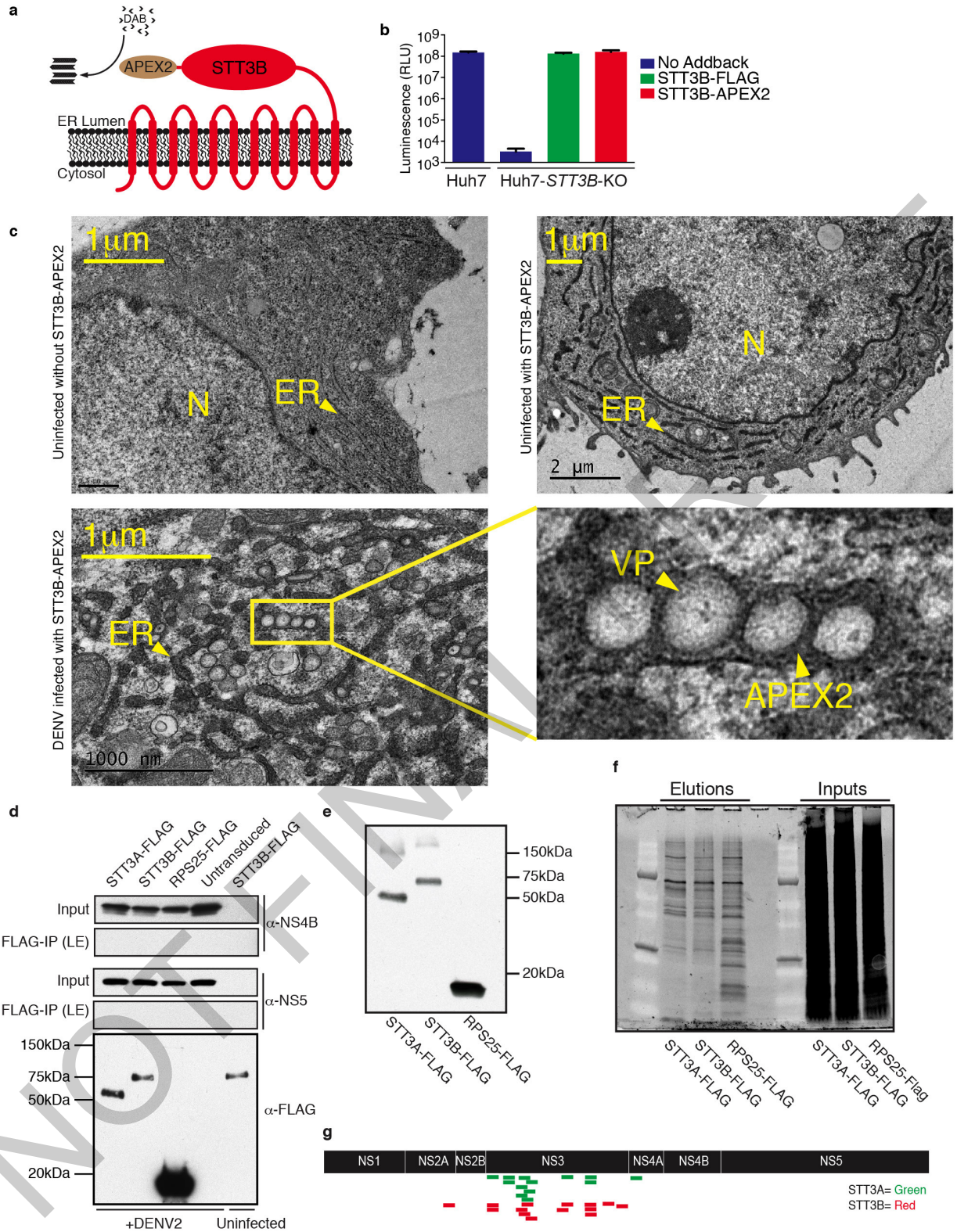
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | Catalytic site mutations introduced in mammalian STT3A and STT3B. a, Catalytic site amino acids highlighted in red as identified in the bacterial STT3 (*Campylobacter lari* pglb). Strong conservation allows their identification in other species. Alignments of STT3 isoforms across different species highlight the conserved catalytic sites that were mutated. The table specifies the amino acid position and the specific triple mutations that were made to abolish catalytic activity. **b**, Huh7 *STT3A* and *STT3B* knockout cells expressing Flag-tagged STT3A and STT3B wild-type and catalytic mutants.

NOT FINAL PROOF

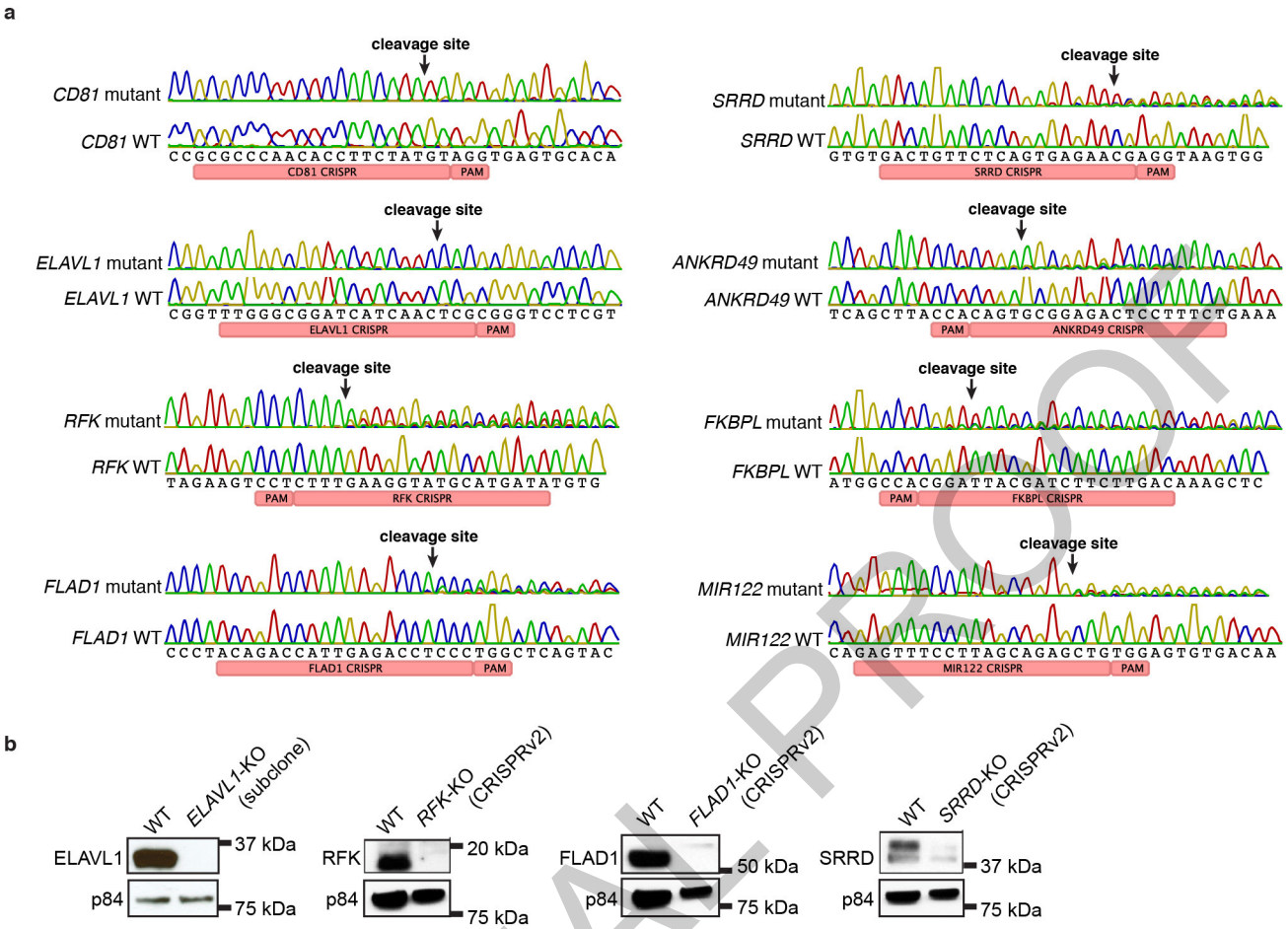
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 6 | Physical interaction between the OST complex and the replication complex of DENV. a, APEX2, a protein tag for electron microscopy was fused to the C terminus of STT3B enabling the imaging of subcellular protein localization by deposition of a polymer of 3,3′-diaminobenzidine (DAB). **b**, Luminescence of Huh7 *STT3B*-knockout cells complemented with STT3B–APEX2 and infected with DENV expressing *Renilla* luciferase. Data are mean and s.d. for triplicate infections. **c**, STT3B localizes on ER membranes in the vicinity of DENV-induced vesicle packets as shown by transmission electron microscopy micrograph of DENV-infected or uninfected Huh7 cells expressing the STT3B–APEX2 construct. N represents the cell nucleus and the arrowheads in samples transfected with STT3B–APEX2 represent APEX polymerized DAB staining in the lumen of the ER or around DENV-induced vesicle packets (VP). **d**, Co-immunoprecipitations of STT3A–Flag and STT3B–Flag from DENV-infected cell lysates. LE, long exposure. **e**, Anti-Flag western blots of immunoprecipitation elutions of DENV-infected cells stably expressing Flag-tagged STT3A, STT3B and RPS25. **f**, SYPRO Ruby staining of elutions and inputs of immunoprecipitations of DENV-infected cell lysates. **g**, Co-immunoprecipitation elutions of DENV-infected lysates were analysed by mass spectrometry and DENV-specific peptides aligned to DENV polyprotein.

NOT FINAL

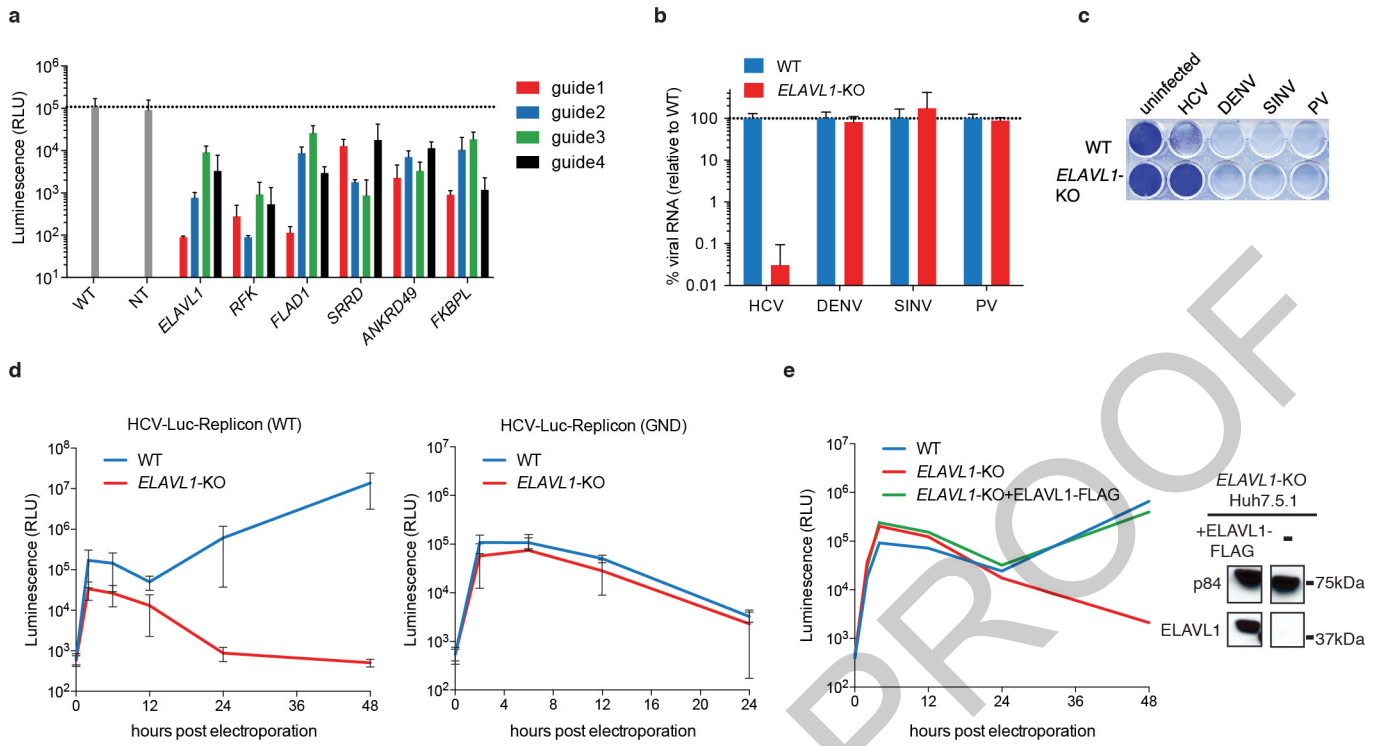
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | Analysis of HCV host factor knockout cell lines. a, Genotyping of CRISPR-induced knockout Huh7.5.1 cells by Sanger sequencing showing the mutated locus and the wild-type reference. CRISPR/Cas9 induces mutations close to the PAM site resulting in frameshifts. *CD81* and *ELAVL1* knockout cell lines are subclones, whereas others are populations of cells mutagenized with lentiCRISPRv2. **b**, Immunoblots of CRISPR-induced knockout cells.

NOT FINAL PROOF

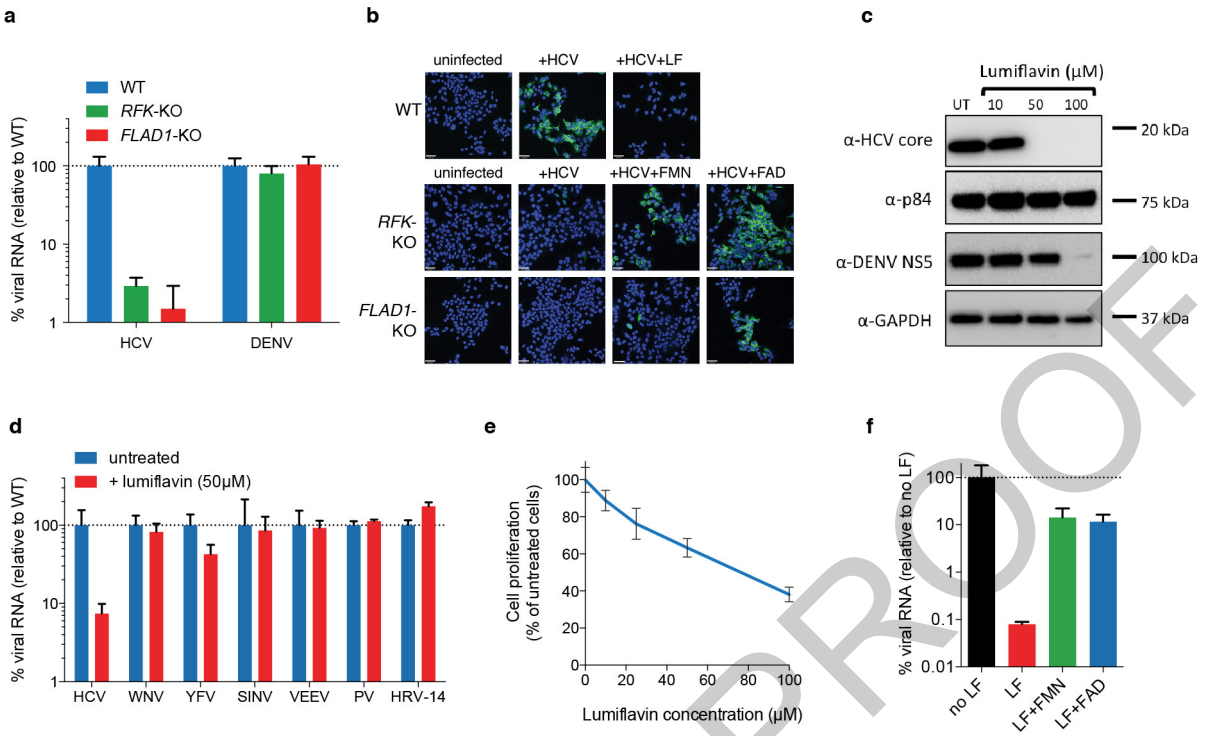
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | ELAVL1 is a critical host factor for HCV replication. a, HCV luciferase infection in knockout cell lines using four different guide RNAs per gene. NT, non-targeting guide RNA. **b**, qPCR of viral RNA in wild-type or ELAVL1-knockout Huh7.5.1 cells. **c**, Crystal violet assay for different RNA virus infections. **d**, HCV replicon assays using wild-type sgJFH1 (left) or GND sgJFH1 replicon. **e**, Transfection of ectopically expressed ELAVL1 restores HCV replication. Western blot of ELAVL1–Flag transfected and untransfected Huh7.5.1 ELAVL1-knockout cells. Data are mean and s.e.m. (qPCR) or s.d. (FFU, RLU) for triplicate infections, except in panel **e**, which was a single infection.

NOT FINAL PROOF

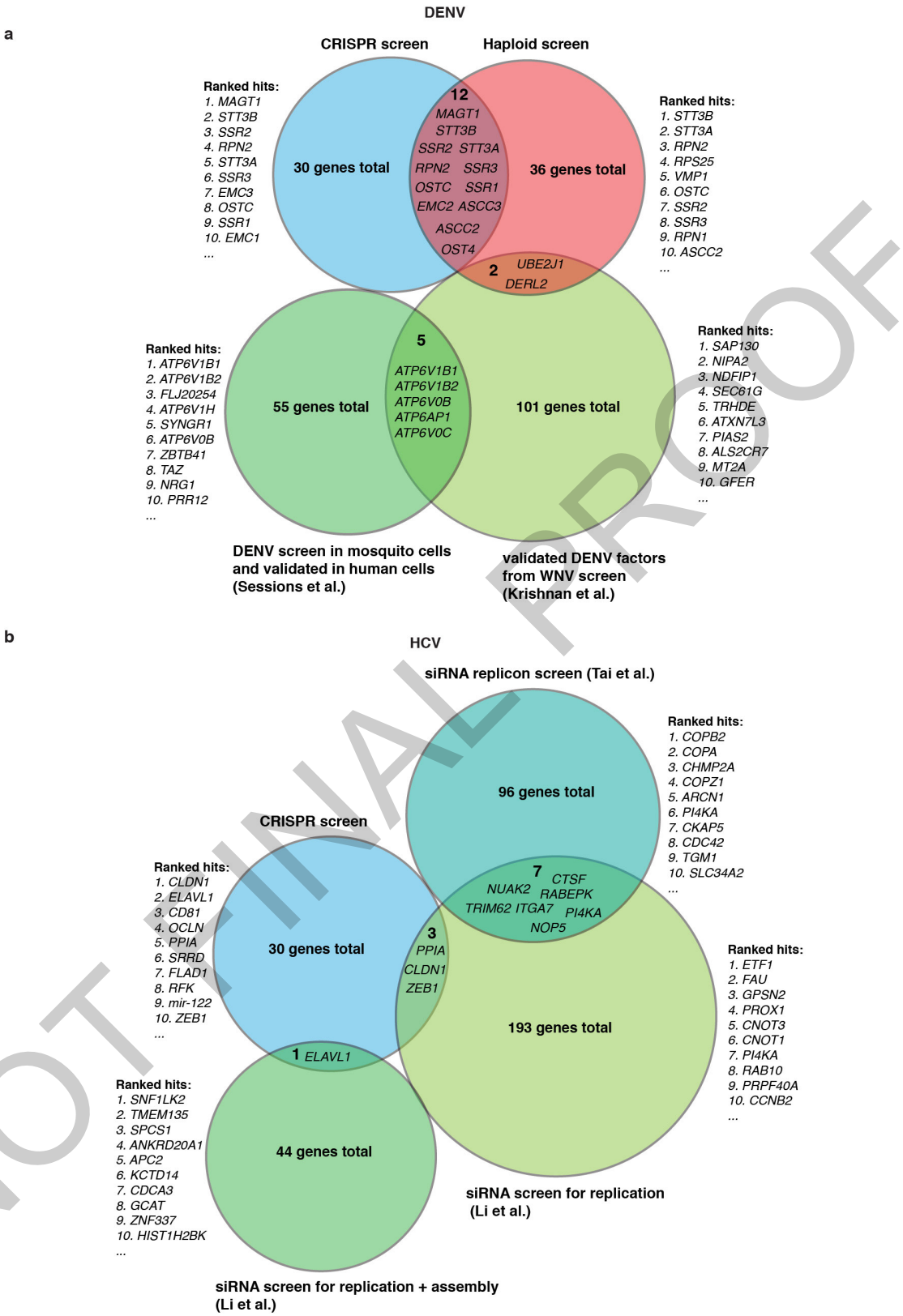
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Lumiflavin inhibits the replication of HCV but not of other RNA viruses. a, qPCR of HCV or DENV RNA replication in wild-type, *RFK*-knockout or *FLAD1*-knockout Huh7.5.1 cells. **b**, Immunofluorescence of HCV infection in wild-type, *RFK*-knockout and *FLAD1*-knockout Huh7.5.1 cells under treatment with lumiflavin, FMN or FAD. HCV core protein (green). Blue denotes DAPI (nuclear) staining. Scale bar, 57 μm. **c**, Western blot for HCV core and DENV NS5 in untreated (UT) and lumiflavin-treated Huh7.5.1 cells. p84 and GAPDH served as loading controls. **d**, qPCR of RNA viruses in untreated or lumiflavin-treated Huh7.5.1 cells. **e**, MTT cell proliferation assay for lumiflavin-treated Huh7.5.1 cells. **f**, Restoration of HCV replication in lumiflavin-treated cells by exogenous addition of FMN or FAD. Data are mean and s.e.m. (qPCR) or s.d. (MTT) for triplicate infections/treatments.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 10 | Comparison of knockout screen results to previous siRNA screens. a, Venn diagram comparing the hits from the CRISPR and haploid screens for DENV host factors to previous siRNA screens from Sessions *et al.*[27] (from supplementary table 2) and Krishnan *et al.*[11] (from supplementary table 1). The top ten validated host factors (by strength of phenotype in the validation screen) for each screen are shown next to the circle. **b**, Venn diagram comparing the hits from the CRISPR screen for HCV host factors to previous siRNA screens from Tai *et al.*[29] (from table S2) and Li *et al.*[28] (from dataset S1).

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18625

A CRISPR screen defines a signal peptide processing pathway required by flaviviruses

Rong Zhang[1], Jonathan J. Miner[1], Matthew J. Gorman[1], Keiko Rausch[2], Holly Ramage[2], James P. White[1], Adam Zuiani[1], Ping Zhang[1,3], Estefania Fernandez[1], Qiang Zhang[1], Kimberly A. Dowd[4], Theodore C. Pierson[4], Sara Cherry[2] & Michael S. Diamond[1,5,6,7]

Flaviviruses infect hundreds of millions of people annually, and no antiviral therapy is available[1,2]. We performed a genome-wide CRISPR/Cas9-based screen to identify host genes that, when edited, resulted in reduced flavivirus infection. Here, we validated nine human genes required for flavivirus infectivity, and these were associated with endoplasmic reticulum functions including translocation, protein degradation, and *N*-linked glycosylation. In particular, a subset of endoplasmic reticulum-associated signal peptidase complex (SPCS) proteins was necessary for proper cleavage of the flavivirus structural proteins (prM and E) and secretion of viral particles. Loss of SPCS1 expression resulted in markedly reduced yield of all Flaviviridae family members tested (West Nile, Dengue, Zika, yellow fever, Japanese encephalitis, and hepatitis C viruses), but had little impact on alphavirus, bunyavirus, or rhabdovirus infection or the surface expression or secretion of diverse host proteins. We found that SPCS1 dependence could be bypassed by replacing the native prM protein leader sequences with a class I major histocompatibility complex (MHC) antigen leader sequence. Thus, SPCS1, either directly or indirectly via its interactions with unknown host proteins, preferentially promotes the processing of specific protein cargo, and Flaviviridae have a unique dependence on this signal peptide processing pathway. SPCS1 and other signal processing pathway members could represent pharmacological targets for inhibiting infection by the expanding number of flaviviruses of medical concern.

We performed a genome-wide inhibition of West Nile virus (WNV)-induced cell death screen using the CRISPR/Cas9 system[3–7] and lentiviruses targeting 19,050 genes (Extended Data Fig. 1a). Whereas in the absence of lentivirus transduction cells did not survive WNV infection, colonies of lentivirus-transduced cells survived; single guide RNAs (sgRNAs) were amplified by PCR and sequenced. We identified 12 genes that were statistically enriched using MAGeCK[8] (Supplementary Tables 1, 2). All 12 genes were endoplasmic reticulum-associated with annotated functions of carbohydrate modification, protein translocation and signal peptide processing, protein degradation, and heat shock response (Fig. 1a).

In validation studies, editing of nine genes resulted in reduced WNV antigen expression following infection of 293T or HeLa cells (Fig. 1a, b) without causing cytotoxicity (Extended Data Fig. 1b). We confirmed the efficiency of gene editing for the proteins for which we could obtain validated antibodies (Extended Data Fig. 1c). Validated genes were tested for effects on related flaviviruses: Zika (ZIKV), Japanese encephalitis (JEV), Dengue serotype 2 (DENV-2), and yellow fever (YFV) viruses. Editing of six of these genes reduced infection by all four flaviviruses (Fig. 1c–f). Editing of *STT3A*, *SEC63*, *SPCS1*, or *SPCS3* resulted in decreased yields of WNV and JEV (Fig. 1g, h). We observed less impact on unrelated positive- or negative-sense RNA viruses (Extended Data Fig. 1d).

As pathogenic flaviviruses are transmitted by arthropods, we evaluated the roles of orthologues of these genes in insect cells. Silencing of *Drosophila* orthologues reduced infection by WNV and DENV-2 (Fig. 2a, b) without appreciably affecting cell viability (Fig. 2c). Decreased WNV infection was also observed in mosquito cells after gene silencing (Fig. 2d). Depletion of *Spase22-23* (orthologue of *SPCS3*) in adult *Drosophila* led to decreased WNV titres (Fig. 2e) and flies heterozygous for *Spase12* (orthologue of *SPCS1*) showed reduced WNV infection (Fig. 2f). Overall, flavivirus infectivity in human and insect cells was dependent on analogous endoplasmic reticulum-associated genes.

Trans-complementation of gene-edited human cells with wild-type alleles rescued flavivirus infectivity (Extended Data Fig. 1e–g). Since we identified the genes encoding two (SPCS1 and SPCS3) of the five components of the Signal Peptidase Complex[9,10], and found that insect SPCS genes also affected flavivirus infection, we focused our study on these genes. Gene silencing in human cells confirmed that SPCS genes were required for optimal flavivirus but not alphavirus infection (Extended Data Fig. 2 and data not shown).

We screened for clonal *SPCS1* and *SPCS3* knockout cells lines. Although we were unable to obtain *SPCS3*$^{-/-}$ clonal lines, *SPCS1*$^{-/-}$ 293T or Huh7.5 cell clones grew, with both alleles containing nonsense deletions (Fig. 3a and Extended Data Fig. 3). WNV, DENV, JEV, YFV, and ZIKV failed to accumulate in the supernatants of *SPCS1*$^{-/-}$ 293T cells (Fig. 3c–f), and WNV infectivity was restored in *trans*-complemented cells (Fig. 3h). However, *SPCS1*$^{-/-}$ cells supported infection by alphaviruses, bunyaviruses, and rhabdoviruses (Fig. 3i–k and Extended Data Fig. 3a). To corroborate these findings, we tested *SPCS1*$^{-/-}$ Huh7.5 cells and found reduced infection by WNV, ZIKV, JEV, and the related *Flaviviridae* member, hepatitis C virus (Extended Data Fig. 3e, f). In comparison, gene editing of the remaining SPCS genes, *SEC11A* and *SEC11C*, had minimal effects on infection (Extended Data Fig. 4).

To determine whether SPCS1 was required for viral translation, replication, or both, we used wild-type and loss-of-function[11] flavivirus replicons encoding reporter genes[12] (Fig. 3b and Extended Data Fig. 5). Transfection of control cells with replicon RNA resulted in low levels of reporter gene activity over the first several hours, which reflects translation of input viral RNA, whereas subsequent signal increases are due to RNA replication. In *SPCS1*$^{-/-}$ cells, high levels of reporter gene expression were observed, indicating that viral RNA translation and replication remained largely intact.

We speculated that SPCS subunits, directly or indirectly, might regulate cleavage of the flavivirus polyprotein[13]. Flavivirus structural (prM and E) and non-structural (NS1 and NS4B) proteins are cleaved by unknown endoplasmic reticulum host signal peptidase(s) (Fig. 3l

[1]Department of Medicine, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. [2]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. [3]Department of Immunology, Institute of Human Virology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China. [4]Viral Pathogenesis Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [5]Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. [6]Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. [7]The Center for Human Immunology and Immunotherapy Programs, Washington University School of Medicine, Saint Louis, Missouri 63110, USA.

© 2016 Macmillan Publishers Limited. All rights reserved

NOT FINAL PROOF

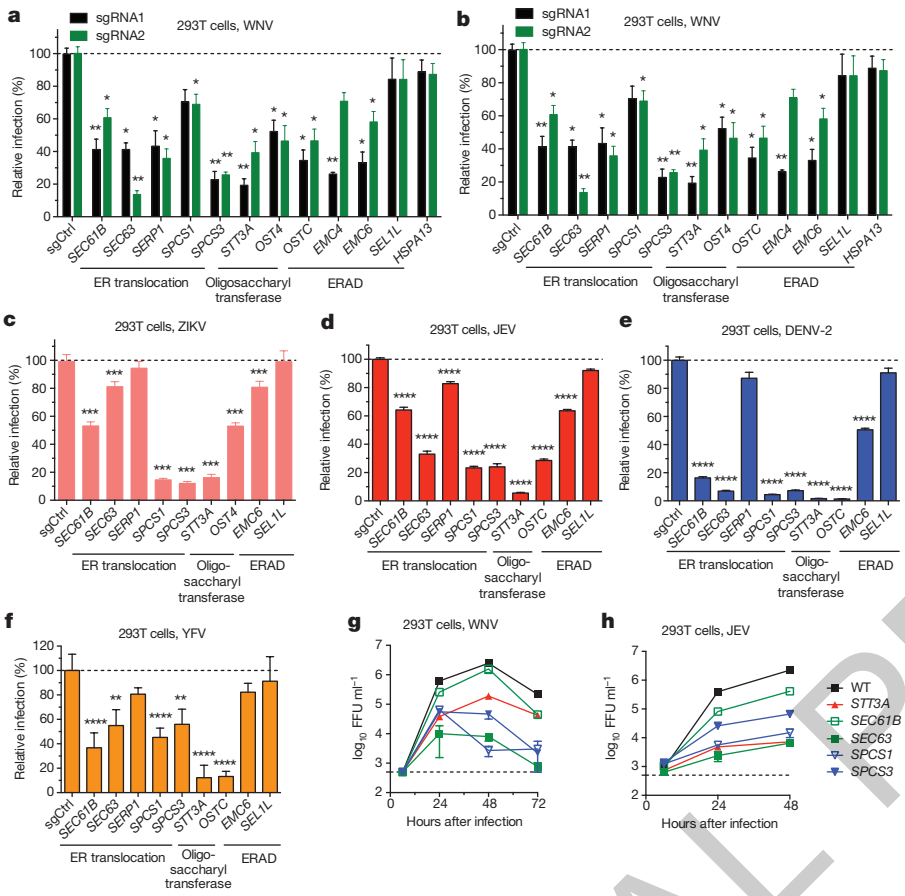


Figure 1 | Genes required for flavivirus infection. a, **b**, Genes were selected for validation based on statistical analysis (Supplementary Tables 1 and 2). Gene-edited 293T (**a**) and HeLa cells (**b**) were infected with WNV and analysed 12 h later for E protein. **c**–**f**, Effect of gene editing on ZIKV (**c**), JEV (**d**), DENV-2 (**e**), and YFV (**f**) infection in 293T cells. The results are the average of two or three independent experiments. **g**, **h**, 293T cells expressing indicated sgRNAs were infected with WNV (**g**) or JEV (**h**) and virus yield was determined. One of two independent experiments performed in triplicate is shown. Statistical significance was determined by ANOVA with a multiple comparisons correction (*$P < 0.05$, **$P < 0.01$, ***$P < 0.0001$; **a**–**f**). Error bars indicate s.e.m. ER, endoplasmic reticulum; ERAD, endoplasmic reticulum-associated degradation.

and refs 14, 15). Gene-edited 293T cells were infected with WNV or JEV, and lysates were analysed. Reduced levels of E and prM proteins were found in *SPCS1*$^{-/-}$ clones and *SPCS1* or *SPCS3* bulk gene-edited cells 12 h after infection, and by 24 h higher molecular mass bands reacted with anti-E or anti-prM/E antibodies[16] (Fig. 3m, n and Extended Data Figs 3g, 6a, b). We next examined whether SPCS1 is required for cleavage of the viral non-structural proteins NS1–NS2A, 2K–NS4B, or NS2B–NS3. In *SPCS1*$^{-/-}$ cells, infection with WNV

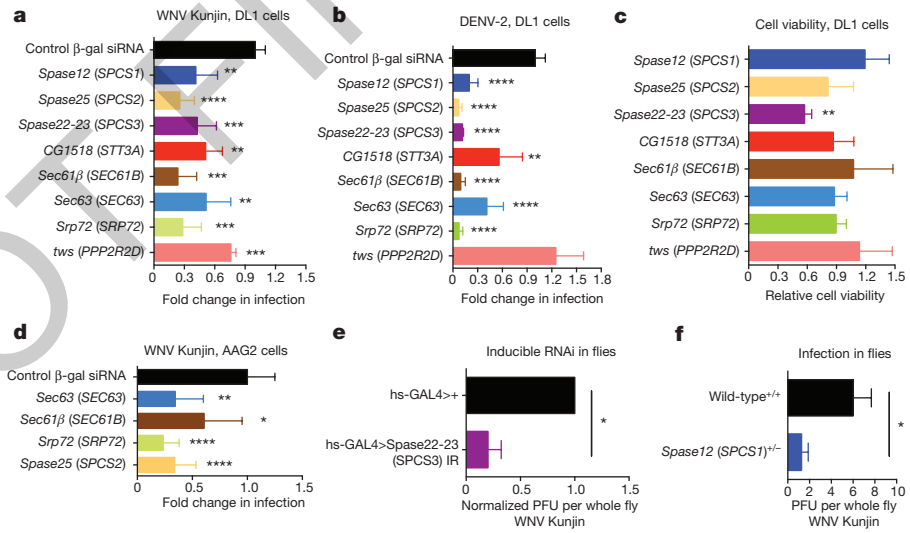


Figure 2 | Endoplasmic reticulum-associated genes are required for flavivirus infection of insect cells. a, **b**, *Drosophila* DL1 cells were treated with dsRNA and infected with WNV (Kunjin) (**a**) or DENV-2 (**b**) for 30 h. Gene names of human orthologues are given in parentheses. The percentage of infected cells was normalized to the control β-galactosidase dsRNA. The data are expressed as the mean normalized value ± s.d. Statistically significant differences were determined by Student's *t*-test (**$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$) and were compared to control dsRNA. The data are pooled from four experiments in duplicate. **c**, Cell viability. DL1 cells were treated with dsRNA and processed 30 h later. **d**, *Aedes agypti* AAG2 cells were treated with dsRNA, infected with WNV (Kunjin) for 30 h, and processed for viral antigens. **e**. *SPCS1* silenced *Drosophila* (Hs-Gal4>UAS–Spase22-23 IR (inverted repeat)) or sibling controls were infected with WNV (Kunjin) and titres measured 7 days later. The fold-change in titres of pools of ten flies from three experiments is shown (normalized mean ± s.d., *$P < 0.05$ by Student's *t*-test). **f**, Wild-type or *Spase12(EY10774)*$^{+/-}$ sibling flies were infected with WNV (Kunjin) and titres measured 7 days later. Data from pools of five flies in three independent experiments is shown (*$P < 0.05$ by Student's *t*-test).

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

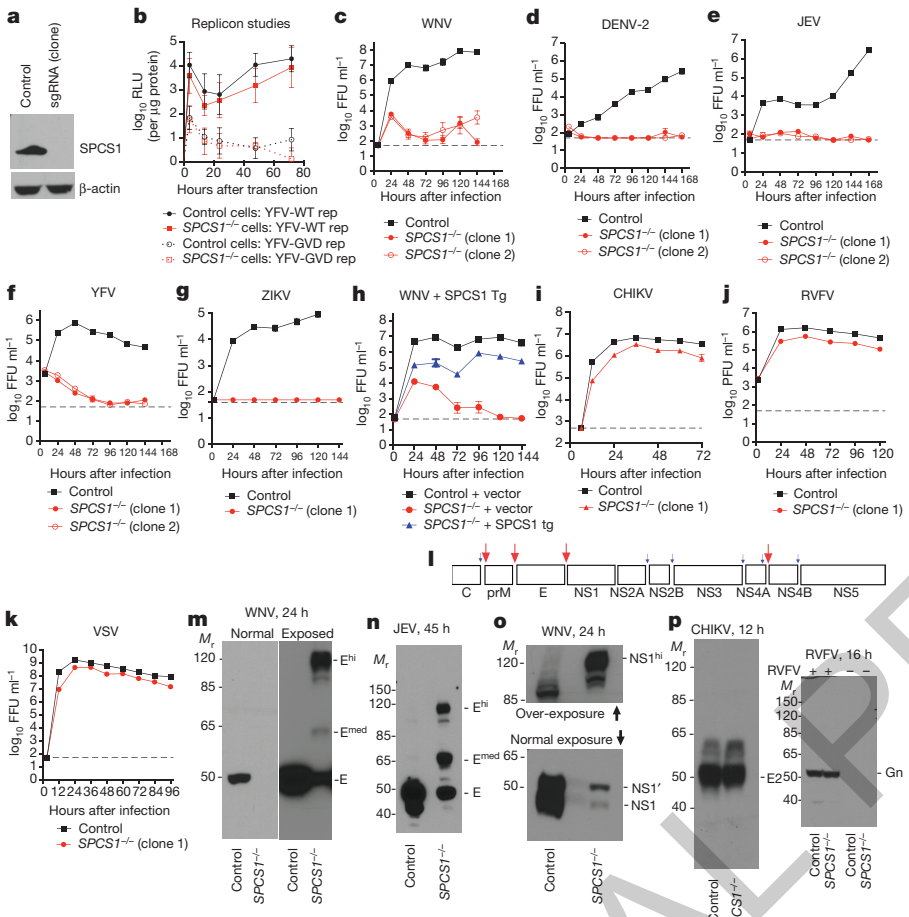


Figure 3 | SPCS1 is required for flavivirus protein processing and infection. a, Western blotting of *SPCS1*$^{-/-}$ 293T cells. **b**, Cells were transfected with YFV–luciferase replicon RNA (wild-type GDD or loss-of-function GVD). Firefly luciferase activity was measured and normalized to intracellular protein levels. The data reflect the average of two or three independent experiments performed in duplicate. **c**–**h**, Cells were infected with WNV (**c**, **h**), DENV-2 (**d**), JEV (**e**), YFV (**f**) or ZIKV (**g**), and viral yield measured. In **h**, cells were *trans*-complemented with an *SPCS1* or control plasmid. Results are the average of two or three independent experiments performed in triplicate. **i**–**k**, Cells were infected with CHIKV (alphavirus), RVFV (bunyavirus), or VSV (rhabdovirus) and viral yield was measured. Results are the average of two or three independent experiments performed in triplicate. **l**, The polyprotein processing strategy of flaviviruses[13]. Red and blue arrows indicate sites of cleavage by host and viral (NS2B–NS3) proteases, respectively. **m**–**o**, Control or *SPCS1*$^{-/-}$ 293T (**m**, **o**) or Huh7.5 (**n**) cells were infected with WNV (**m**, **o**) or JEV (**n**). Lysates were blotted with (**m**) anti-WNV E, (**n**) anti-JEV E, or (**o**) anti-WNV NS1 monoclonal antibodies. Higher molecular mass bands (E^{hi}, E^{med}, and $NS1^{hi}$) that react with anti-flavivirus monoclonal antibodies are indicated. One experiment of three is shown. **p**, 293T cells were infected with CHIKV or RVFV. Lysates were blotted with anti-CHIKV E2 or anti-RVFV Gn monoclonal antibodies. One experiment of two is shown. For gel source data, see Supplementary Fig. 1. FFU, fluorescence-focus forming unit; PFU, plaque-forming unit.

resulted in decreased expression of NS1 and the accumulation of higher molecular mass bands (Fig. 3o). We detected lower levels of NS4B protein in *SPCS1*$^{-/-}$ cells; in transfection studies with a tagged 2K–NS4B plasmid, a higher molecular mass band was observed. For NS1–NS2A and NS3, we did not detect aberrant cleavage (Extended Data Fig. 6). We also tested the effects on HCV E2 glycoprotein and found decreased levels in *SPCS1*$^{-/-}$ cells (Extended Data Fig. 7). In comparison, alphavirus or bunyavirus glycoproteins, which also require endoplasmic reticulum processing[17,18], showed intact expression in *SPCS1*$^{-/-}$ cells (Fig. 3p and Extended Data Fig. 3b, c).

To isolate the effects of the SPCS complex from infection, we transfected a prM–E plasmid, which produces subviral particles (SVPs)[19]. Immunoblotting of cell lysates for E and prM proteins showed reduced levels and higher molecular mass bands in *SPCS1*- or *SPCS3*-deficient cells, and these changes correlated with a reduction in the number of SVPs (Extended Data Fig. 8a–c). We tested whether cleavage of flavivirus protein signal sequences depended on SPCS1. We transfected WNV structural (capsid (*C*), *prM*, *M*, *E*) and secreted non-structural (*NS1*) genes with native or MHC class I (K^b) signal sequences into *SPCS1*$^{-/-}$ cells, and evaluated protein expression (Fig. 4).

Expression of C protein from a C–prM–E plasmid was equivalent in control and *SPCS1*$^{-/-}$ cells, although in the absence of the viral protease, C did not migrate at its normal size (Extended Data Fig. 8d). However, cleavage of the downstream proteins prM and E was reduced in *SPCS1*$^{-/-}$ cells. When NS2B–NS3 was supplied *in trans*, C was cleaved from prM–E and accumulated at the correct size in control and *SPCS1*$^{-/-}$ cells. Thus, expression or cleavage of C is not affected by SPCS1.

We next evaluated expression of prM and M. When the native *prM* leader sequence was used, expression of prM and its furin-cleavage product M was reduced in *SPCS1*$^{-/-}$ cells (Fig. 4a, groups 1 and 3). Substitution of the K^b leader rescued prM and M expression in *SPCS1*$^{-/-}$ cells only when *prM* was on a separate plasmid (Fig. 4a, group 2) but not as a prM–E plasmid (Fig. 4a, group 4). Thus, specific leader sequences determine the dependence of prM and M protein expression on SPCS1, and downstream proteins can modulate processing efficiency.

When *E* was transfected, its expression was largely independent of SPCS1 or the K^b leader sequence (Fig. 4b, groups 1 and 2). When *E* was cloned downstream of prM, accumulation of E was not detected in *SPCS1*$^{-/-}$ cells (Fig. 4b, groups 3 and 4). This finding suggested that the native leader sequence of *E* was not cleaved in *SPCS1*$^{-/-}$ cells when presented as an 'internal' leader sequence or that epistatic effects of the upstream prM protein reduced the stability of E protein. To test which of these possible explanations was correct, we performed ^{35}S pulse-chase studies in prM–E-transfected cells. In control cells, only a single E protein band was visible, indicating rapid prM–E cleavage. However, prM–E and E bands were both present in *SPCS1*$^{-/-}$ cells (Fig. 4c, top) and remained in an endoplasmic reticulum-resident form (Fig. 4c, bottom). A short 3-min ^{35}S pulse showed a delay in the cleavage of prM–E in *SPCS1*$^{-/-}$ cells (Fig. 4d).

We assessed the expression of NS1, which also requires endoplasmic reticulum-dependent signal sequence cleavage. When *NS1* was transfected into cells, SPCS1 was not required for expression (Fig. 4e, group 1). When *NS1* was cloned downstream of *E* (Fig. 4e, groups 2 and 3) or prM–E (Fig. 4e, groups 4 and 5), NS1 levels were reduced in *SPCS1*$^{-/-}$ cells. After blotting with an anti-NS1 monoclonal antibody, a 90-kDa band was visible in blots from *SPCS1*$^{-/-}$ cells (Fig. 4e, group 2), which probably represented uncleaved E–NS1; this result was corroborated by blotting for E protein (Fig. 4b, groups 5 and 6). Thus, placement of the NS1 leader sequence into an internal position rendered it more dependent on SPCS1 for cleavage.

Flavivirus SVPs can be produced after transfection of *prM* and *E* on single or separate plasmids[20,21]. Transfection of prM–E encoding native

FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

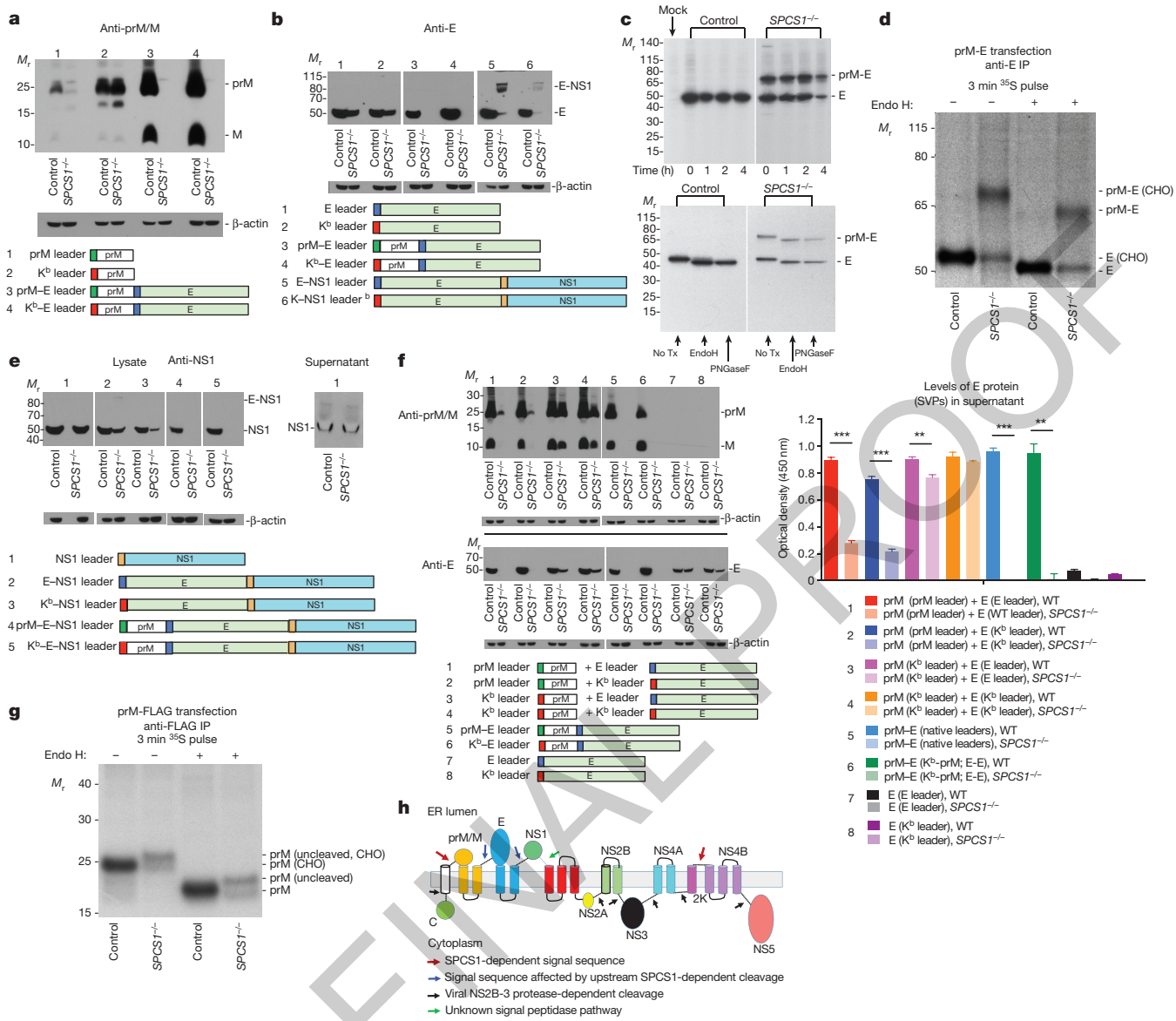


Figure 4 | SPCS1 is required for cleavage of the C-prM leader peptide and internal leader sequences. a, 293T cells were transfected with prM or prM–E plasmids containing native (C–prM (green box) and prM–E (blue box)) or K^b (red box) leader sequences. Blots of lysates were probed with anti-prM/M monoclonal antibodies. One experiment of three is shown. **b**, 293T cells were transfected with E, prM–E, and E–NS1 plasmids containing native leader sequences (as in **a** and E–NS1 (orange box)) or a K^b leader. Blots of lysates were probed with anti-E monoclonal antibodies. A higher molecular mass band corresponds to uncleaved E–NS1. One experiment of two is shown. **c**, **d**, 293T cells were transfected with a prM–E plasmid. After 24 h, cells were labelled for 40 min (**c**) or 3 min (**d**) with ^{35}S cysteine-methionine. Lysates were immunoprecipitated with an anti-E protein monoclonal antibody before SDS–PAGE. **c**, Top, cysteine–methionine was added for chase times (0–4 h). **c** (bottom) and **d**, Immunoprecipitates were untreated or treated with Endo H or PNGase F for 1 h at 37 °C. One experiment of two is shown. **e**, 293T cells were transfected with NS1, E–NS1, or prM–E–NS1 plasmids containing native viral or K^b leaders. Blots of lysates or supernatants were probed with anti-NS1 monoclonal antibodies. One experiment of two is shown. **f**, 293T cells were transfected with prM + E, prM–E, or E plasmids containing viral or K^b leaders. Left, blots of lysates were probed with anti-prM/M or anti-E monoclonal antibodies. One experiment of two is shown. Right, levels of SVPs in supernatant at 24 h. Data are pooled from independent experiments performed in triplicate (**$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$; unpaired t-test). **g**, 293T cells were transfected with prM–Flag. After 24 h, cells were labelled for 3 min with ^{35}S cysteine–methionine. Lysates were immunoprecipitated with anti-Flag protein monoclonal antibodies. **h**, Model of processing of flavivirus structural and non-structural proteins based on infection and transfection studies and the literature[14,15]. Arrows indicate cleavage sites requiring SPCS1, sites affected by upstream SPCS1-dependent events, sites cleaved by the viral NS2B–NS3 protease, and sites cleaved via an SPCS1-independent pathway. For gel source data, see Supplementary Fig. 1.

or K^b and native internal signal sequences resulted in loss of expression of prM and E or SVPs in *SPCS1*$^{-/-}$ cells (Fig. 4f, groups 5 and 6). When *prM* and *E* were co-transfected, the proteins were detected in *SPCS1*$^{-/-}$ cell lysates (Fig. 4f, groups 1 and 2) and supernatant, albeit at lower levels. In *SPCS1*$^{-/-}$ cells, prM negatively affected E but not NS1 production (Fig. 4f (compare groups 1, 2, and 7) and Extended Data Fig. 8e), possibly because of its chaperone-like function for E protein[20]. Defects in co-expression of prM and E in *SPCS1*$^{-/-}$ cells were corrected by inserting the K^b leader sequence in front of the *prM* gene (Fig. 4f, *groups 3 and 4*). A 3-min ^{35}S pulse and immunoprecipitation experiment in SPCS1$^{-/-}$ cells showed an uncleaved form of prM (Fig. 4g).

To assess whether host surface proteins require SPCS1 for signal peptide processing, we profiled *SPCS1*$^{-/-}$ Jurkat T cells. Whereas ten antigens showed no difference in surface expression, levels of CD49d–CD29, ULBP1, and HLA-E were reduced by two-to-threefold (Extended Data Fig. 9a–c). A decrease in surface expression of ULBP1

© 2016 Macmillan Publishers Limited. All rights reserved

NOT FINAL PROOF

has been reported in cells deficient in SPCS1 or SPCS2 expression[22], although this phenotype was not explored. In an unbiased approach, we analysed the secretome of *SPCS1*$^{-/-}$ 293T cells by mass spectrometry. Of the approximately 245 secreted proteins identified, only 35 were downregulated in *SPCS1*$^{-/-}$ cells, and the fold-changes were small (Extended Data Fig. 10 and Supplementary Table 4). We validated 3 of 5 as being reduced in supernatants of *SPCS1*$^{-/-}$ cells (Supplementary Table 5). Despite profound effects on flavivirus protein processing, an absence of SPCS1 only modestly affected the expression of host proteins.

The differential requirement of SPCS1 for viral and host protein processing suggests that components of the SPCS complex in mammalian and probably insect cells facilitate the cleavage of particular signal peptides in specific contexts. There may be additional requirements for some viruses, as interactions between SPCS1 and the HCV NS2 and E2 proteins have been reported[23].

A recent study performed an analogous CRISPR-based screen with WNV[24]. Endoplasmic reticulum-associated genes were identified that prevented WNV-induced cell death. We identified three of these genes (*EMC4*, *EMC6*, and *SEL1L*), as did an siRNA screen[25]. Virtually all human gene 'hits' identified in our screen had insect orthologues required for optimal flavivirus infection. A subset of our genes were also identified in RNAi screens in *Drosophila* cells[26,27]. The endoplasmic reticulum is a focal site in the flavivirus lifecycle because it supports translation, polyprotein processing, replication, and virion morphogenesis. The identification of host gene targets that are selectively required for flavivirus infection but not cell survival provides intriguing candidates for pharmacological inhibition.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 November 2015; accepted 6 June 2016.
Published online 17 June 2016.

1. Diamond, M. S. & Pierson, T. C. Molecular insight into Dengue virus pathogenesis and its implications for disease control. *Cell* **162,** 488–492 (2015).
2. Weaver, S. C. *et al.* Zika virus: History, emergence, biology, and prospects for control. *Antiviral Res.* **130,** 69–80 (2016).
3. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
4. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2,** e00471 (2013).
5. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343,** 84–87 (2014).
6. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, Mdel. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32,** 267–273 (2014).
7. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343,** 80–84 (2014).
8. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15,** 554 (2014).
9. Evans, E. A., Gilmore, R. & Blobel, G. Purification of microsomal signal peptidase as a complex. *Proc. Natl Acad. Sci. USA* **83,** 581–585 (1986).
10. Meyer, H. A. & Hartmann, E. The yeast SPC22/23 homolog Spc3p is essential for signal peptidase activity. *J. Biol. Chem.* **272,** 13159–13164 (1997).
11. Khromykh, A. A., Kenney, M. T. & Westaway, E. G. *trans*-Complementation of flavivirus RNA polymerase gene NS5 by using Kunjin virus replicon-expressing BHK cells. *J. Virol.* **72,** 7270–7279 (1998).
12. Jones, C. T., Patkar, C. G. & Kuhn, R. J. Construction and applications of yellow fever virus replicons. *Virology* **331,** 247–259 (2005).
13. Lindenbach, B. D., Murray, C. L., Thiel, H. J. & Rice, C. M. in *Fields Virology* Vol. 1 (eds Knipe, D. M. & Howley, P. M.) 712–746 (Lippincott Williams & Wilkins, 2013).
14. Chambers, T. J., Grakoui, A. & Rice, C. M. Processing of the yellow fever virus nonstructural polyprotein: a catalytically active NS3 proteinase domain and NS2B are required for cleavages at dibasic sites. *J. Virol.* **65,** 6042–6050 (1991).
15. Falgout, B., Pethel, M., Zhang, Y. M. & Lai, C. J. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J. Virol.* **65,** 2467–2475 (1991).
16. Throsby, M. *et al.* Isolation and characterization of human monoclonal antibodies from individuals infected with West Nile Virus. *J. Virol.* **80,** 6982–6992 (2006).
17. Barth, B. U., Wahlberg, J. M. & Garoff, H. The oligomerization reaction of the Semliki Forest virus membrane protein subunits. *J. Cell Biol.* **128,** 283–291 (1995).
18. Löber, C., Anheier, B., Lindow, S., Klenk, H. D. & Feldmann, H. The Hantaan virus glycoprotein precursor is cleaved at the conserved pentapeptide WAASA. *Virology* **289,** 224–229 (2001).
19. Schalich, J. *et al.* Recombinant subviral particles from tick-borne encephalitis virus are fusogenic and provide a model system for studying flavivirus envelope glycoprotein functions. *J. Virol.* **70,** 4549–4557 (1996).
20. Lorenz, I. C., Allison, S. L., Heinz, F. X. & Helenius, A. Folding and dimerization of tick-borne encephalitis virus envelope proteins prM and E in the endoplasmic reticulum. *J. Virol.* **76,** 5480–5491 (2002).
21. Hanna, S. L. *et al.* N-linked glycosylation of west nile virus envelope proteins influences particle assembly and infectivity. *J. Virol.* **79,** 13262–13274 (2005).
22. Gowen, B. G. *et al.* A forward genetic screen reveals novel independent regulators of ULBP1, an activating ligand for natural killer cells. *eLife* **4,** e08474 (2015).
23. Suzuki, R. *et al.* Signal peptidase complex subunit 1 participates in the assembly of hepatitis C virus through an interaction with E2 and NS2. *PLOS Pathog.* **9,** e1003589 (2013).
24. Ma, H. *et al.* A CRISPR-based screen identifies genes essential for West-Nile-Virus-induced cell death. *Cell Rep.* **12,** 673–683 (2015).
25. Krishnan, M. N. *et al.* RNA interference screen for human genes associated with West Nile virus infection. *Nature* **455,** 242–245 (2008).
26. Yasunaga, A. *et al.* Genome-wide RNAi screen identifies broadly-acting host factors that inhibit arbovirus infection. *PLOS Pathog.* **10,** e1003914 (2014).
27. Sessions, O. M. *et al.* Discovery of insect and human dengue virus host factors. *Nature* **458,** 1047–1050 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by NIH grants U19 AI083019 (M.S.D.), U19 AI106772 (M.S.D.), R01 AI104972 (M.S.D.), and T32 AI007163 (E.F.) and by the Washington University Institute of Clinical and Translational Sciences (UL1 TR000448 from the National Center for Advancing Translational Sciences and P41 GM103422-35 from the National Institute of General Medical Sciences). T.C.P and K.A.D are supported by the intramural program of NIAID. We thank R. Kuhn, A. Garcia-Sastre, H. Zhao, D. Fremont, X. Wang, and R. Townsend for reagents, experimental advice, and data analysis; P. Erdmann-Gilmore, R. Connors, Y. Mi, and H. Lin for expert technical assistance; and X. de Lamballerie and the European Virus Archive goes Global (EVAg) for consenting to the use of H/PF/2013 ZIKV strain for this study under a material transfer agreement with the EVAg parter, Aix-Marseille Université.

Author Contributions R.Z. performed the primary CRISPR/Cas9 screen and viral protein transfection experiments. Validation studies in cells with different viruses were performed by R.Z., P.Z., M.J.G, A.Z., E.F., and M.S.D. J.J.M. performed the pulse-chase experiments and mass cytometry. J.P.W. performed studies with replicons. K.R. and H.R performed the insect cell and insect experiments. Q.Z. performed the secretome analysis and mass spectrometry. R.Z., S.C., T.C.P., and M.S.D. designed the experiments. K.A.D. and T.C.P. provided key reagents. Q.Z. performed data analysis. S.C. and M.S.D. wrote the initial draft of the manuscript, with the other authors contributing to editing into the final form.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S.D. (diamond@borcim.wustl.edu).

Reviewer Information *Nature* thanks W. Wei and the other anonymous reviewer(s) for their contribution to the peer review of this work.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Cells and viruses. Vero, BHK21, HeLa, U2OS, Huh7.5, and 293T cells were cultured at 37 °C in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS). C6/36 *Aedes albopictus* cells were cultured at 28 °C in L15 supplemented with 10% FBS and 25 mM HEPES pH 7.3. *Drosophila* DL1 cells were cultured at 28 °C in Schneider's medium supplemented with 10% FBS as described[28]. Jurkat cells were cultured at 37 °C in RPMI 1640 supplemented with 10% FBS and 10 mM HEPES pH 7.3. All cell lines were originally acquired from American Type Culture Collection or colleagues (Huh7.5) and were tested and judged free of mycoplasma contamination. The following viruses were used in screening and validation studies: WNV (New York 2000), WNV (Kunjin), JEV (14-14-2 vaccine and Bennett strains), DENV-2 (16681 and New Guinea C strains), ZIKV (H/PF/2013), YFV (17D vaccine), CHIKV (2006 La Reunion OPY1), LACV (original strain), VSV (Indiana), HCV (J6/JFH), and SINV (Toto). With the exception of HCV (see below), all other viruses were propagated in BHK21, Vero, or C6/36 cells and titrated by standard plaque or focus-forming assays[29].

Viral growth analysis. 293T or Huh7.5 cells were infected with WNV (multiplicity of infection (MOI) 0.01), JEV (14-14-2 strain, MOI 0.05 or 0.5; Bennett strain, MOI 0.05), DENV-2 (MOI 3), YFV (MOI 1), ZIKV (MOI 0.05), CHIKV (MOI 0.01), SINV (MOI 0.01), RVFV (MOI 1), or VSV (MOI 0.01). After 2 h of incubation, cells were washed three times and samples were titrated on Vero cells. For HCV growth analysis, control and SPCS1 gene-edited Huh7.5 cells were inoculated at an MOI of 1 with virus derived from a growth-adapted JFH-1 infectious clone[30]. Cells were rinsed 6 h after infection to remove unbound virus and samples were collected every 24 h for 7 days. Viral titres in the supernatant were quantified by focus-forming assay, as described previously[31].

Pooled sgRNA screen and data analysis. A pooled library encompassing 122,411 different sgRNAs against 19,050 human genes was derived by the Zhang laboratory[32] and obtained from a commercial source (Addgene). The library was packaged using a lentivirus expression system and 293T cells were transfected using FugeneHD (Promega). Forty-eight hours after transfection, supernatants were collected, clarified by centrifugation (3,500 rpm × 20 min), filtered, and aliquotted for storage at −80 °C.

For the screen, we generated clonal 293T-Cas9 cells by transfecting the lentiCas9-Blast plasmid (Addgene 52962) using FugeneHD transfection reagent (Promega), blasticidin selection, and limiting dilution. These 293T-Cas9 cells were transduced with lentiviruses encoding individual sgRNAs at an MOI of 0.3. After selection with puromycin for 7 days, $\sim 2 \times 10^8$ cells were infected with WNV (MOI of 1) and then incubated for 2–3 weeks. In parallel, untransduced 293T-Cas9 cells were infected to ensure virus-induced infection and cell death. The experiments were performed in parallel as either duplicate or triplicate technical replicates in two independent biological experiments.

Genomic DNA was extracted from the uninfected cells (5×10^7) or the cells (3×10^7) that survived WNV infection, and sgRNA sequences were amplified[5], and subjected to next generation sequencing using an Illumina HiSeq 2500 platform. The sgRNA sequences against specific genes were recovered after removal of the tag sequences using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and cutadapt 1.8.1.

Gene validation. The cut-off for candidate gene 'hits' was made using a published computational tool (MAGeCK)[8] and reflected sequencing data showing multiple different sgRNAs per gene, the number of sequencing reads per gene, and the enrichment of a given sgRNA compared to the uninfected cell library (Supplementary Tables 1, 2). From this, we identified 12 genes that showed statistically significant enrichment compared to uninfected cells. These candidate genes were tested for validation by using 3–5 independent sgRNAs per gene from the parent library and cloning them into the plasmid pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene plasmid 62988). The control sgRNAs were used from the parent library. Plasmids were transfected into 293T or HeLa cells using FugeneHD transfection reagent and puromycin was added one day later. Three days later, puromycin was removed, and cells were allowed to recover for three additional days before infection with different viruses.

For flow cytometric analyses, gene-edited 293T cells were infected with WNV (MOI 5, 12 h), JEV (MOI 50, 22 h), ZIKV (MOI 10, 30 h), DENV-2 (MOI 3, 32 h), YFV (MOI 3, 38 h), CHIKV (MOI 1, 6 h), SINV (MOI 10, 6 h), LACV (MOI 5, 6 h), or VSV-GFP (MOI 1, 5.5 h). Gene-edited HeLa cells were infected with WNV (MOI 3, 24 h). At the indicated times, cells were fixed with 1% paraformaldehyde (PFA) diluted in PBS for 20 min at room temperature and permeabilized with Perm buffer (HBSS (Invitrogen), 10 mM HEPES, 0.1% (w/v) saponin (Sigma), and 0.025% NaN_3 (Sigma)) for 10 min at room temperature. Cells then were rinsed one additional time with Perm buffer, transferred to a U-bottom plate, and incubated for 1 h at 4 °C with 1 μg ml^{-1} of the following virus-specific antibodies: WNV (human E16 (ref. 33)); DENV2 (mouse E18 (ref. 34)); JEV (mouse E18 (ref. 34)); YFV (mouse E60 (ref. 34)); CHIKV (CHK-11 (ref. 35)); SINV (ascites fluid, ATCC VR-1248AF), LACV (807-31 and 807-33, gift from A. Pekosz). After washing, cells were incubated with an Alexa Fluor 647-conjugated goat anti-mouse or anti-human IgG (Invitrogen) for 1 h at 4 °C. Cells were fixed in 1% PFA in PBS, processed on a FACS Array (BD Biosciences), and analysed using FlowJo software (Tree Star).

Validation also was performed by an infectious virus yield assay. Bulk gene-edited 293T cells were infected with WNV (MOI 0.01) or JEV (MOI 0.5). Supernatants were collected at specific times after infection and focus-forming assays were performed in 96-well plates as described previously[36]. Following infection, cell monolayers were overlaid with 100 μl per well of medium (1 × DMEM, 4% FBS) containing 1% carboxymethylcellulose, and incubated for 22 h (WNV) or 36 h (JEV) at 37 °C with 5% CO_2. Cells were then fixed by adding 100 μl per well of 1% PFA directly onto the overlay at room temperature for 40 min. Cells were washed twice with PBS, permeabilized (in 1 × PBS, 0.1% saponin, and 0.1% BSA) for 20 min, and incubated with antibodies specific for WNV (humanized E16 (ref. 33)) or JEV (mouse E18 (ref. 34)) E glycoprotein for 1 h at room temperature. After being rinsed twice, cells were incubated with species-specific HRP-conjugated secondary antibodies (Sigma). After further washing, foci were developed by incubating in 50 μl per well of TrueBlue peroxidase substrate (KPL) for 10 min at room temperature, after which time cells were washed twice in water. Well images were captured using Immuno Capture software (Cell Technology Ltd), and foci counted using BioSpot software (Cell Technology Ltd).

Insect cell and fly infections. dsRNAs were generated as described[37]. To silence genes using RNAi, insect cells were passaged into serum-free medium containing dsRNAs targeting the indicated genes. Cells were serum-starved for 1 h, after which complete medium was added and cells were incubated for 3 days. Cells were infected with WNV (Kunjin strain) at an MOI of 4 or DENV-2 (NGC strain) at an MOI of 1 for 30 h and then processed for microscopy with automated image analysis as described[38]. Control (hs>+) or *Spcs3*-depleted (hs>Spase22-23 IR (Bloomington)) 4–7-day-old female flies were heat shocked (37 °C) for 1 h for three consecutive days to deplete the gene of interest and challenged with WNV (Kunjin) (5 PFU). At day 7 after infection, pools of 10 flies were crushed and titred by plaque assay. Three independent experiments were performed. Heterozygous flies (*Spase12(EY10774)*) were outcrossed to wild-type flies and either wild-type or *Spase12* heterozygous sibling controls were challenged with Kunjin for 7 days and groups of 5 flies were titred.

siRNA treatments in human cells. Human U2OS cells were transfected with siRNAs against control, *SPCS1*, *SPCS2*, or *SPCS3* genes for three days, infected with WNV (Kunjin strain) or DENV (MOI 1) for 18 h, and then processed for microscopy with automated image analysis as described[38].

Replicon transfection and analysis. Two types of replicons were used.

SP6-generated YFV replicons. The wild-type and NS5 polymerase mutant (GDD→GVD) YFV replicons (YF-FFLuc2A, wild-type and GVD) have been published previously[12] and were a gift from R. Kuhn. Capped replicon RNA was generated using SP6 polymerase with an mMESSAGE mMACHINE kit according to the manufacturer's instructions (Thermo Fisher Scientific). RNA was purified using an RNEasy kit (Qiagen) and 2 μg was transfected into control or *SPCS1*$^{-/-}$ Huh7.5 cells using Lipofectamine 3000 according to the manufacturer's instructions (Thermo Fisher Scientific). At specified times, cells were collected, lysed, and processed for firefly luciferase activity using a commercial kit (Promega). Cleared lysates were tested for Fluc activity using the Dual-Luciferase Reporter Assay System (Promega) and the protein concentration was quantified using a BCA assay kit (ThermoFisher). Fluc activity (relative light units, RLU) was normalized by subtracting background luminescence of transfected cells collected at the time of transfection, then the adjusted RLU was divided by the total protein content (in μg) to yield RLU per μg protein.

cDNA launched WNV replicons. The construction of wild-type and NS5 polymerase mutant (GDD→GVD) WNV replicons (lineage I, strain New York 1999) was based on a previously described cDNA launched molecular clone system[39]. The backbone of this strategy, a plasmid containing a truncated WNV genome under the control of a CMV promoter (pWNV-backbone), was designed to be complemented via ligation of a structural gene DNA fragment; transfection of pWNV-backbone alone does not result in production of a self-replicating RNA molecule. Using overlap extension PCR and unique restriction endonuclease sites, pWNV-backbone was modified by the introduction of a fragment downstream of the CMV promoter encoding [5′UTR-cyclization sequence of capsid-FMDV2a protease-signal sequence of E-NS1] to complement the NS2→NS5-3′UTR already present in the pWNV-backbone plasmid, generating the replicon plasmid pWNVI-rep. The reporter gene *GFP* was then cloned upstream of the FMDV2a protease sequence via a unique MluI site to generate pWNVI-rep–GFP. The construction and organization of this WNV lineage I replicon is analogous to a previously described lineage II WNV replicon (pWNVIIrep–GFP)[40]. Finally, QuikChange mutagenesis (Agilent Technologies) was used to delete the enhancer portion of the CMV immediate early enhancer/promoter, generating pWNVI-minCMV-rep–GFP, and to generate the

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

GDD→GVD NS5 polymerase variant. Although the CMV enhancer–promoter combination commonly found in cloning vectors results in robust and constitutive expression, inclusion of only the minimal CMV promoter (no enhancer) results in low-level expression[41]. As such, direct transfection of pWNVI-minCMV-rep–GFP results in a dim GFP signal, which reflects translation of the RNA generated by DNA-dependent RNA translation. RNA polymerase-dependent replication of the wild-type (but not GVD mutant) replicon results in higher production of GFP over time. The eGFP is bracketed by the FMDV2a autocleavage site, and does not rely on host or viral proteases for processing. Wild-type and NS5 GVD variants of pWNVI-minCMV-rep–GFP (200 ng) were transfected into 10^4 control or gene-edited 293T cells (96-well plates) using Lipofectamine 2000. At various times after transfection, cells were collected, cooled to 4 °C, stained sequentially with a biotinylated anti-9NS1 (ref. 42) (or biotin anti-CHIKV negative control monoclonal antibodies) and Alexa 647-conjugated streptavidin. In some samples, cells were fixed with 4% PFA in PBS (10 min, room temperature) and permeabilized with 0.1% (w/v) saponin. Cells were processed for two-colour flow cytometry using a MACs Quant Analyzer 10 (Miltenyi Biotec).

Plasmid transfections. 293T gene-edited cells were transfected with the following genes that were derived from a WNV infectious cDNA clone[43] and then cloned into a pHLsec backbone (gift from D. Fremont): V5–C–prM–E, prM, prM–Flag (3 × Flag), E, prM–E, prM–E–NS1, E–NS1, NS1, NS1–NS2A–Flag (includes full-length NS1 and 231 amino acids of NS2A fused to a C-terminal 3 × Flag), and 2K–NS4B–haemagglutinin tag (HA). These plasmids were obtained from colleagues (e.g., 2K-NS4B-HA[44], gift from A. Garcia-Sastre) or in some cases engineered to contain either native WNV signal sequences (C-prM, 18 amino acids beyond the C terminus of C; prM-E, 17 C-terminal amino acids of prM; E-NS1, 24 C-terminal amino acids of E) or the signal sequence of mouse K^b class I MHC (N-terminal 21 amino acids). Plasmids were transfected into gene-edited 293T cells using FugeneHD reagent (Promega) according to the manufacturer's instructions. Supernatants containing prM-E subviral particles (SVPs) were collected 24 h after transfection, filtered through a 0.2-μm filter, and stored aliquotted at −80 °C. For the capture ELISA, Nunc MaxiSorp polystyrene 96-well plates were coated overnight at 4 °C with mouse E60 monoclonal antibodies[34] (5 μg ml^{-1}) in a pH 9.3 carbonate buffer. Plates were washed three times in enzyme-linked immunosorbent assay (ELISA) wash buffer (PBS with 0.02% Tween 20) and blocked for 1 h at 37 °C with ELISA block buffer (PBS, 2% bovine serum albumin, and 0.02% Tween 20). Supernatants from prM-E plasmid transfected cells were captured on plates coated with E60 for 90 min at room temperature. Subsequently, plates were rinsed five times in wash buffer and then incubated with humanized anti-WNV E16 (1 μg ml^{-1} in block buffer) for 1 h at room temperature. Plates were washed five times and then incubated with pre-absorbed biotinylated goat anti-human IgG antibody (1 μg ml^{-1}; Jackson Laboratories) for 1 h at room temperature in blocking buffer. Plates were washed again five times and then sequentially incubated with 2 μg ml^{-1} of horseradish peroxidase-conjugated streptavidin (Vector Laboratories) and tetramethylbenzidine substrate (Dako). The reaction was stopped with the addition of 2 N H_2SO_4 to the medium, and emission (450 nm) was read using an iMark microplate reader (Bio-Rad).

Western blotting. For virus infected samples, cells were infected with WNV (MOI 200–1,000, 24 h), JEV (MOI 150, 45 h), CHIKV (MOI 5, 12 h), SINV (MOI 5, 16 h), RVFV (MOI 2.5, 16 h), or HCV (MOI 5, 48 or 72 h). Cells (10^6) were lysed directly in 30 μl RIPA buffer (Cell Signaling) with 0.1% SDS and a cocktail of protease inhibitors (Sigma-Aldrich). Samples were prepared in LDS buffer (Life Technologies) under non-reducing or reducing (dithiothreitol) conditions. After heating (70 °C, 10 min), samples were electrophoresed using 7% Tris-Acetate or 4–12%, 10% or 12% Bis-Tris gels (Life Technologies) and proteins were transferred to PVDF membranes using an iBlot2 Dry Blotting System (Life Technologies). Membranes were blocked with 5% non-fat dry powdered milk and probed with antibodies against SPCS1 (11847-1-AP, Proteintech), SPCS2 (14872-1-AP, Proteintech), SPCS3 (ab91222, Abcam), SEC11A (14753-1-AP, Proteintech), SEC11C (HPA026816, Sigma) and SEC61B (ab15576, Abcam). For studies with prM–E, prM, E, NS1, NS1–2A–Flag, or 2K–NS4B–Flag-transfected or virus-infected cells, membranes were probed with anti-E (human E16; mouse CHK-48[35]; mouse anti-JEV, oligoclonal pool), anti-NS1 (mouse 8-NS1), anti-NS3 (W1018-54, USBio), anti-NS4B (rabbit polyclonal antibody[45], gift from W.I. Lipkin) anti-prM (human CR4293[16] or rabbit WNV-M (IMG-5099A, IMGENEX)), anti-Flag (F1804, Sigma), and the relevant secondary antibodies. For validation of the secretome experiments, supernatants were electrophoresed and PVDF membranes were probed with anti-CXCL16 (ab101404, Abcam), anti-SFRP1 (ab126613, Abcam), anti-RNASET2 (ab169655, Abcam), anti-LGALS3BP(ab81489, Abcam),anti-SLITL2 (ab173758, Abcam), anti-PEDF (ab157207, Abcam), anti-NPC2 (19888-1-AP, Proteintech), anti-CREG1 (12220-1-AP, Proteintech), and the relevant secondary antibodies. Western blots were developed using SuperSignal West Pico Chemiluminescent Substrate or SuperSignal West Femto Maximum Sensitivity Substrate (Life Technologies).

Metabolic labelling, pulse-chase, and immunoprecipitation experiments. Pulse- and pulse-chase experiments were performed as described previously[46]. After starvation in methionine/cysteine-free DMEM for 30 min, 293T cells were labelled metabolically with 300 or 500 μCi ml^{-1} [^{35}S]-methionine/cysteine (PerkinElmer Life Sciences) at 37 °C for 3 or 40 min. Cells then were washed three times in PBS and immediately lysed or incubated in DMEM supplemented with non-radiolabelled cysteine (500 μg ml^{-1}) and methionine (100 μg ml^{-1}). Cells lysis was performed in 400 μl of 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM PMSF, 1 mM EDTA, 5 μg ml^{-1} aprotinin, 5 μg ml^{-1} leupeptin, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS. After preclearing with an irrelevant human monoclonal antibody protein A-agarose (Thermo Fisher Scientific) complex, lysates were incubated for 1 h at 4 °C with humanized monoclonal E16 and E60 monoclonal antibodies or anti-Flag and then with protein A-agarose for 2 h. The immunoprecipitates were washed seven times in 50 mM Tris-HCl, pH 7.4, 150 mm NaCl, 1 mM PMSF, 1 mM EDTA, 5 μg ml^{-1} aprotinin, 5 μg ml^{-1} leupeptin, 1% Triton X-100, 1% sodium deoxycholate, and 0.1% SDS, and then analysed by SDS–PAGE under reducing conditions, followed by fluorography. Some immunoprecipitates were incubated with 20 mU endoglycosidase H or PNGase F (New England BioLabs) for 1 h at 37 °C before SDS–PAGE and fluorography.

293T cell viability assay. A Vybrant MTT cell viability assay (Life Technologies) was used according to the manufacturer's instructions. Briefly, 10 μl of 12 mM MTT (4,5-dimethylthiazol-2-yl-2-5-diphenyltetrazolium bromide) was added to 10^5 293T cells (different gene-edited lines, with or without WNV infection) in 100 μl phenol-red free medium. Cells were incubated for 4 h at 37 °C, at which time medium was removed and formazan crystals solubilized in 100 μl of DMSO were added for 10 min at 37 °C. Liquid was analysed for absorbance at 540 nm using a Synergy H1 Hybrid Plate Reader (Biotek).

Flow and mass cytometry analysis of Jurkat T cells. The antibodies and conjugates used are listed in Supplementary Table 6. For flow cytometry studies, wild-type and *SPCS1* gene-edited Jurkat T cells were incubated with fluorophore-conjugated monoclonal antibodies for 30 min at 4 °C and then washed three times in PBS containing 5% FBS. Cells were immediately processed on an LSRII flow cytometer and data were analysed using FlowJo 10.0.7. For mass cytometry studies, wild-type and *SPCS1* gene-edited Jurkat T cells were labelled with monoclonal antibodies conjugated with transition element isotopes and analysed on a CyTOF 2 mass cytometer (Fluidigm DVS Sciences). Data were analysed using Cytobank (http://wustl.cytobank.org) and FlowJo 10.0.7.

Secretome analysis of SPCS1$^{-/-}$ 293T cells. Wild-type and *SPCS1*$^{-/-}$ 293T cells were cultured in poly-D-lysine treated flasks in FreeStyle 293 Expression Medium (ThermoFisher) supplemented with 10% FBS. At 90% confluence, cells were washed four times with pre-warmed PBS, then twice with pre-warmed FreeStyle 293 Expression Medium, and maintained in FreeStyle 293 Expression Medium without FBS for 48 h. Supernatants were collected and clarified by centrifugation at 1,000*g* for 5 min, and then 10,000*g* for 20 min at 4 °C. Samples were concentrated with Amicon Ultra-15 Centrifugal Filter Units (Millipore) at 5,000*g* for 1 h in the presence of 1 × protease inhibitors (S8830, Sigma). The concentrates were collected and stored at −80 °C. After thawing on ice, the samples were exchanged twice in digestion buffer (Tris, 0.1 M, pH 8.5 containing 8 M urea) by centrifugation (~4,000*g*, 2 h) in Amicon Ultracel 3K units to a volume of ~100 μl. The solubilized samples were reduced with 2 mM DTT (ThermoScientific) for 30 min at 37 °C followed by alkylation at room temperature for 30 min with 7 mM iodoacetamide (Sigma) in the dark. The alkylated samples were treated with 7 mM DTT for 15 min at room temperature. After dilution, the samples were digested with LysC (1 μg) (Sigma) overnight at 37 °C with agitation (ThermoMixer). After dilution of the samples to 1.5 M urea with Tris buffer, trypsin was added (5 μg) (Sigma) was added and the incubation was continued overnight at 37 °C with mixing. The digested samples were acidified with to a concentration of 1% tri-flouro acetic acid (TFA). The peptides were desalted with a SepPak (50 mg) with 0.1%TFA/70% acetonitrile in an elution volume (2 ml). The lyophilized peptides were quantified with a fluorescent assay (Thermo Fisher) and 2 μg was labelled with TMT-6 reagents according to the vendor. The labelled peptides were desalted and the samples were transferred to PCR tubes (0.5 ml) and positioned in 96-well holders for robotic solid phase extraction (SPE). Each digest was extracted sequentially with one C4 tip (Glygen BIOMEK NT3C04) and one porous graphite carbon micro-tip (Glygen BIOMEK NT3CAR) with the following auto-pipetting steps: (i) wet tips with AcN/FA (60%/1%) (10 × 25 μl); (ii) equilibrate tips with AcN/FA (1%/1%) (10 × 25 μl); (iii) extract peptides with repetitive aspirations of the digest (50 × 25 μl); (iv) wash loaded tips with AcN/FA (1%/1%) (10 × 25 μl); and (v) elute peptides with AcN/FA (60%/1%) (5 × 65 μl). The SPE eluents were pooled and dried in a SpeedVac centrifuge and transferred to an autosampler vial for LC–MS analysis.

The remainder of the peptides were dissolved in the binding buffer (100 mM Tris, pH 7.8 containing NaCl (0.5 M), $MnCl_2$ (1 mM) and $CaCl_2$ (1 mM). The dried lectins (Con-A and WGA) Sigma were dissolved in binding buffer (4 mg ml^{-1}).

© 2016 Macmillan Publishers Limited. All rights reserved

The rCA120 (10 mg ml^{-1}), Con-A and WGA were added to the peptide solution (36 µl and 10 µl, respectively). After incubation at room temperature, the mixture was transferred to a YM-10 Microcon filter unit. After centrifugation (14,000*g*) for 10 min and washing with binding buffer (100 µl), the filter unit was transferred to another tube. The peptides were released with the addition of PNGase (10 units) in 100 µl of ammonium bicarbonate buffer (50 mM) after incubation at 37 °C for 1.5 h. The enzyme addition and incubation was repeated and the peptides recovered with one wash of PNGase buffer. The peptides were acidified to 5% formic acid and desalted, labelled with TMT-6, and prepared for LC–MS as described above for the total pool of peptides.

LC–MS analysis. LC–ESI/MS/MS analysis was conducted with a Q-Exactive Plus mass spectrometer coupled to an EASY-nanoLC 1000 system (Thermo-Fisher). For each Hp-RP fraction, 2 µl of sample was loaded onto a 75 µm i.d. × 25 cm Acclaim PepMap 100 RP column (Thermo-Fisher Scientific). Peptide separations were started with 95% mobile phase A (0.1% FA) for 5 min and increased to 30% B (100% ACN, 0.1% FA) over 180 min, followed by a 25-min gradient to 45% B, a 5-min gradient to 95% B and wash at 90% B for 7 min, with a flow rate of 300 nl min^{-1}. Full-scan mass spectra were acquired by the Orbitrap mass analyser in the mass-to-charge ratio (m/z) of 375–1,400 and with a mass resolving power set to 70,000. Fifteen data-dependent high-energy collisional dissociations were performed with a mass resolving power set to 35,000, a fixed first m/z of 100, an isolation width of 0.7 m/z, and the normalized collision energy (NCE) setting of 32. The maximum injection time was 50 ms for parent-ion analysis and 105 ms for product-ion analysis. Target ions already selected for MS/MS were excluded dynamically for 30 s. An automatic gain control target value of 3×10^6 ions was used for full MS scans and 10^5 ions for MS/MS scans. Peptide ions with charge states of one or greater than six were excluded from MS/MS interrogation.

Protein identification and quantification with TMT. All raw data were processed using Proteome Discoverer (version 2.1.0.81, Thermo-Fischer Scientific). MS/MS spectra were searched with SequestHT engine against the human UniRef database (69,021 entries; version 2014_05), assuming the digestion enzyme was trypsin with a maximum of 2 missed cleavage allowed. The searches were performed with a fragment ion mass tolerance of 0.02 Da and a parent ion tolerance of 20 ppm. Deamidation of asparagine and glutamine, acetylation and TMT 6-plex derivatization of N termini and oxidation of methionine were specified in Proteome Discoverer as variable modifications. Iodoacetamide derivatization of cysteine and TMT 6-plex derivatization of lysine were specified as fixed modifications. Peptide spectral matches (PSM) were validated using percolator based on q-values at a 1% FDR[47]. Peptides were filtered to 1% FDR and grouped into proteins at 1% FDR as specified in Proteome Discoverer. The intensities of TMT reporter ions were determined with Proteome Discoverer at a mass tolerance of 0.01 Da and used for peptide quantifications. The median values of peptide intensities that can be assigned to a same protein was used to represent protein intensities. Peptide identifications that can be assigned to more than one protein were removed from protein quantification

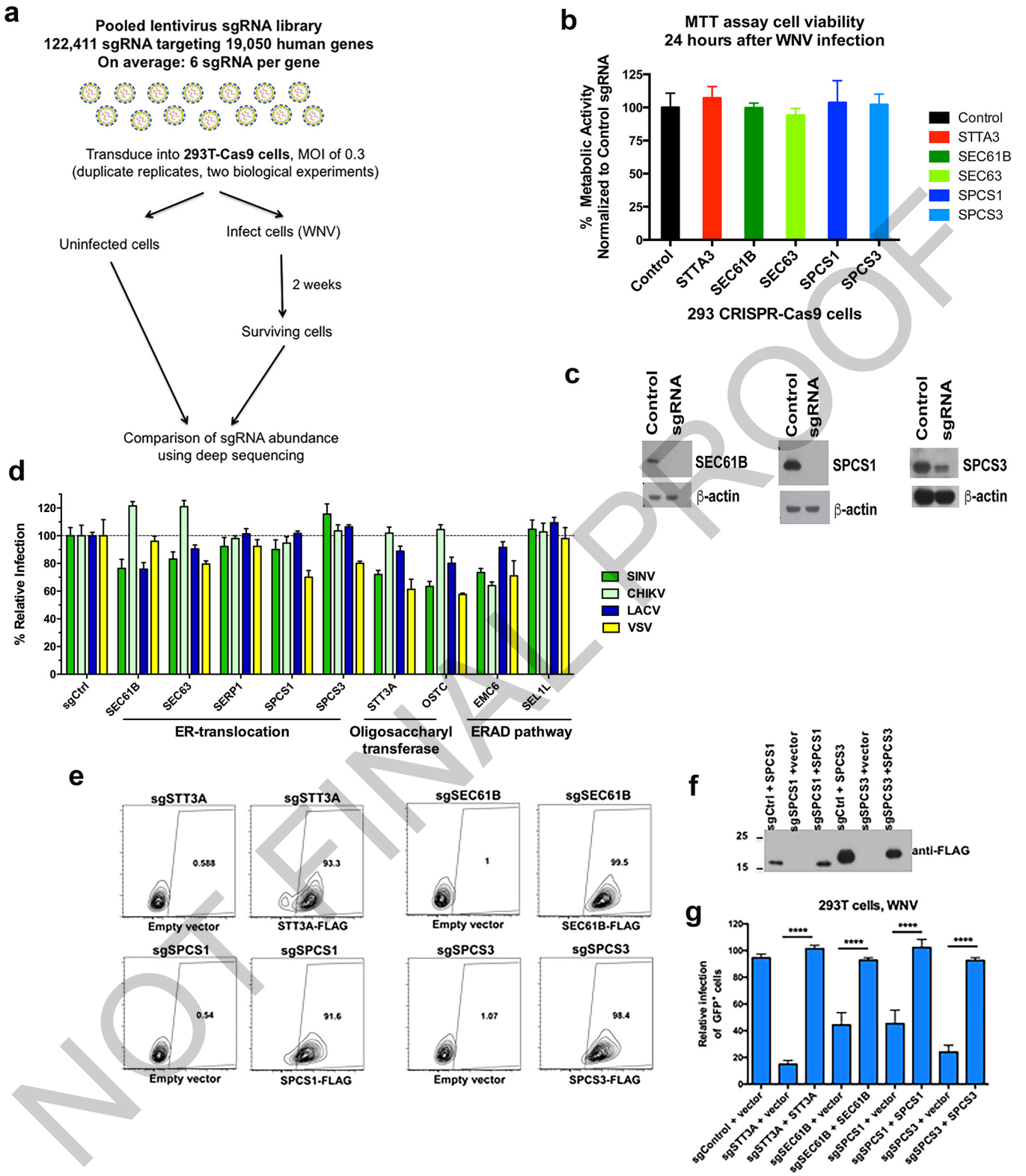
Proteomic Data Analysis. Normalization in protein ratios was applied in that the median ratios are $\log_2 0$. Data analysis was performed with the free software environment for statistical computing and graphics, R (http://www.R-project.org). Gene ontology analysis was carried out using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[48,49]. Data from duplicated LC/MS/MS analysis were first averaged and protein abundance ratios were $\log_2$-transformed before statistical analysis. A one-way ANOVA with Benjamini-Hochberg correction was performed to assess the statistical significance in protein abundance changes between wild type and *SPCS1*$^{-/-}$ cells.

Statistical analysis. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. Statistical significance was assigned when P values were <0.05 using GraphPad Prism Version 5.04. Viral antigen staining after expression of sgRNA was analysed using a one-way ANOVA adjusting for repeated measures with a Dunnett's multiple comparison test or with a Mann–Whitney test depending on the number of comparison groups. Analysis of levels of E protein in the supernatant from CRISPR-Cas9 gene edited cells was analysed by a one-way ANOVA. Analysis of siRNA in insect and human cells was performed using a Student's t-test or ANOVA.

28. Rose, P. P. *et al.* Natural resistance-associated macrophage protein is a cellular receptor for sindbis virus in both insect and mammalian hosts. *Cell Host Microbe* **10,** 97–104 (2011).
29. Brien, J. D., Lazear, H. M. & Diamond, M. S. Propagation, quantification, detection, and storage of West Nile virus. *Curr. Protoc. Microbiol.* **31,** 15D 13 11–15D 13 18 (2013).
30. Jiang, J. & Luo, G. Cell culture-adaptive mutations promote viral protein-protein interactions and morphogenesis of infectious hepatitis C virus. *J. Virol.* **86,** 8987–8997 (2012).
31. Sabo, M. C. *et al.* Neutralizing monoclonal antibodies against Hepatitis C Virus E2 protein bind discontinuous epitopes and inhibit infection at a postattachment step. *J. Virol.* **85,** 7005–7019 (2011).
32. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Meth.* **11,** 783–784 (2014).
33. Oliphant, T. *et al.* Development of a humanized monoclonal antibody with therapeutic potential against West Nile virus. *Nat. Med.* **11,** 522–530 (2005).
34. Oliphant, T. *et al.* Antibody recognition and neutralization determinants on domains I and II of West Nile Virus envelope protein. *J. Virol.* **80,** 12149–12159 (2006).
35. Pal, P. *et al.* Development of a highly protective combination monoclonal antibody therapy against Chikungunya virus. *PLoS Pathog.* **9,** e1003312 (2013).
36. Fuchs, A., Pinto, A. K., Schwaeble, W. J. & Diamond, M. S. The lectin pathway of complement activation contributes to protection from West Nile virus infection. *Virology* **412,** 101–109 (2011).
37. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303,** 832–835 (2004).
38. Hackett, B. A. *et al.* RNASEK is required for internalization of diverse acid-dependent viruses. *Proc. Natl Acad. Sci. USA* **112,** 7797–7802 (2015).
39. Lin, T. Y. *et al.* A novel approach for the rapid mutagenesis and directed evolution of the structural genes of west nile virus. *J. Virol.* **86,** 3501–3512 (2012).
40. Pierson, T. C. *et al.* A rapid and quantitative assay for measuring antibody-mediated neutralization of West Nile virus infection. *Virology* **346,** 53–65 (2006).
41. Mishin, V. P., Cominelli, F. & Yamshchikov, V. F. A 'minimal' approach in design of flavivirus infectious DNA. *Virus Res.* **81,** 113–123 (2001).
42. Chung, K. M. *et al.* Antibodies against West Nile Virus nonstructural protein NS1 prevent lethal infection through Fc gamma receptor-dependent and -independent mechanisms. *J. Virol.* **80,** 1340–1351 (2006).
43. Beasley, D. W. *et al.* Envelope protein glycosylation status influences mouse neuroinvasion phenotype of genetic lineage 1 West Nile virus strains. *J. Virol.* **79,** 8339–8347 (2005).
44. Muñoz-Jordán, J. L. *et al.* Inhibition of alpha/beta interferon signaling by the NS4B protein of flaviviruses. *J. Virol.* **79,** 8004–8013 (2005).
45. Medigeshi, G. R. *et al.* West Nile virus infection activates the unfolded protein response, leading to CHOP induction and apoptosis. *J. Virol.* **81,** 10849–10860 (2007).
46. Miner, J. J. *et al.* Cytoplasmic domain of P-selectin glycoprotein ligand-1 facilitates dimerization and export from the endoplasmic reticulum. *J. Biol. Chem.* **286,** 9577–9586 (2011).
47. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4,** 923–925 (2007).
48. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols* **4,** 44–57 (2009).
49. Huang, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 1 | See next page for caption.

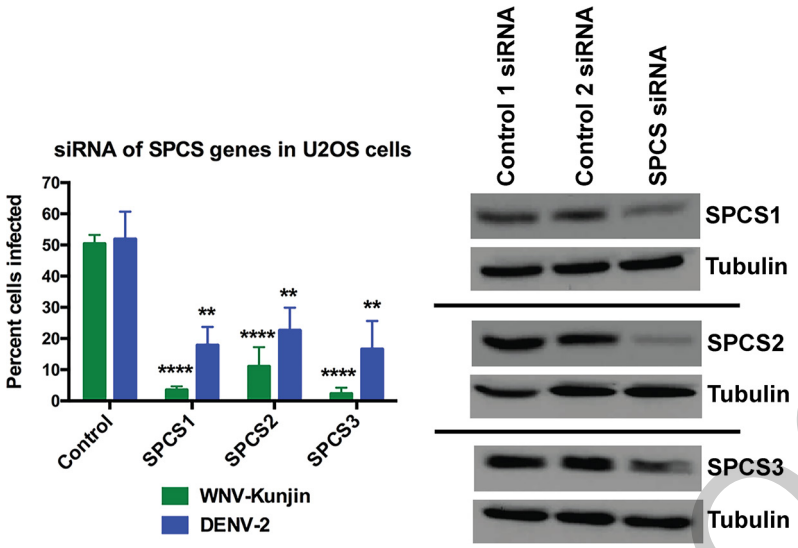
NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Figure 1 | Results of CRISPR-Cas9 screen. a, Scheme of gene-editing screen. **b**, Analysis of cell viability of gene-edited cells. WNV-infected CRISPR-Ca9 edited bulk cells were evaluated for cell viability using a metabolic MTT assay 24 h after infection. The results were pooled from several independent experiments performed in duplicate and data were compared to cells edited with a control sgRNA. None of the differences were statistically significant compared to the control. **c**, Western blotting confirms the efficiency of gene editing of *SEC61B*, *SPCS1*, and *SPCS3*. β-actin is included as a loading control. **d**, Effect of gene editing on infection by other RNA viruses. sgRNA-edited bulk selected cell populations were infected with alphaviruses (SINV or CHIKV), a bunyavirus (LACV) or a rhabdovirus (VSV). Cells were analysed for intracellular viral antigen staining by flow cytometry using virus-specific monoclonal antibodies. The data are representative of two independent experiments and are expressed as relative infection (viral antigen expression) compared to the sgRNA control. **d**–**f**, *Trans*-complementation of sgRNA gene-edited cells with Flag-tagged genes. **d**, Individual sgRNA bulk gene-edited cell lines were *trans*-complemented with cDNA expressing C-terminal Flag-tagged versions of their respective genes and GFP or an empty vector control and GFP. Transfected cells were analysed by flow cytometry for expression of the Flag-tag in the GFP^+ cells. The data are representative of two independent experiments. **e**, Western blotting of *SPCS1* and *SPCS3 trans*-complemented genes after incubation with an anti-Flag antibody. **f**, Individual sgRNA cell lines were *trans*-complemented with cDNA expressing C-terminal Flag-tagged versions of their respective genes and GFP or an empty vector control and GFP. Transfected cells were infected with WNV (MOI 5) and 12 h later cells were stained for intracellular E antigen and processed by flow cytometry. The data are representative of three independent experiments performed in triplicate and reflect the percentage of WNV-infected cells in the fraction that expressed GFP. The indicated comparisons were statistically significant (****$P < 0.0001$), as determined by the Mann–Whitney test. For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

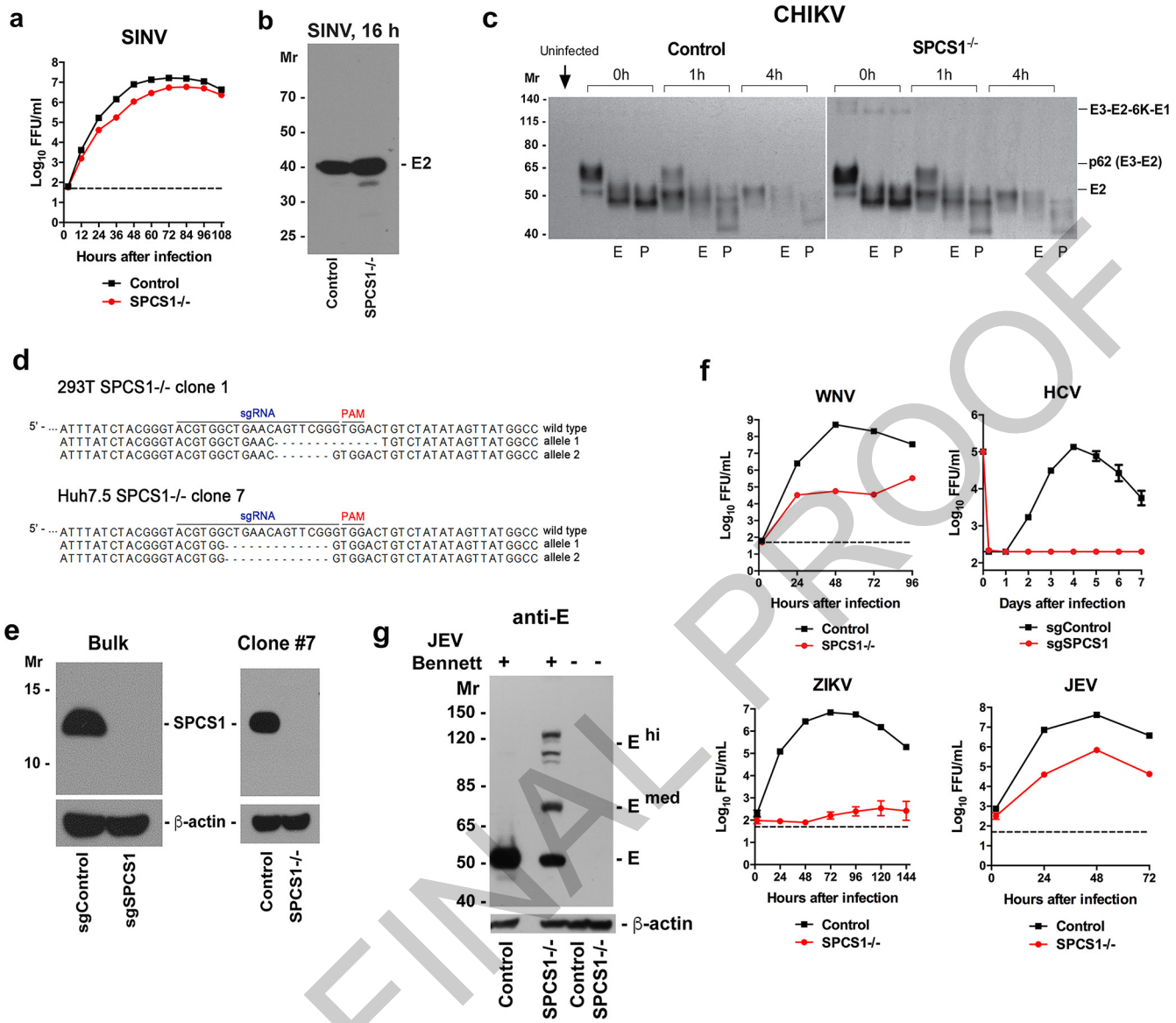
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 2 | Gene silencing of SPCS genes in human U2OS cells. Human U2OS cells were transfected with either control or *SPCS1*, *SPCS2*, or *SPCS3* siRNAs and infected with WNV (Kunjin strain) or DENV (MOI 1) for 18 h. Left, the percentage of infected cells was determined by automated fluorescence microscopy. The data are expressed as the mean normalized value ± s.d. $^{**}P < 0.01$; $^{***}P < 0.0001$ compared to control siRNA by ANOVA with a multiple comparison correction. The data are pooled from three independent experiments assayed in quadruplicate. No reduction in infection of CHIKV or SINV was observed after SPCS gene silencing (data not shown). Right, western blotting of SPCS proteins in gene-silenced U2OS cells. Representative results are shown and tubulin is included as a loading control. For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

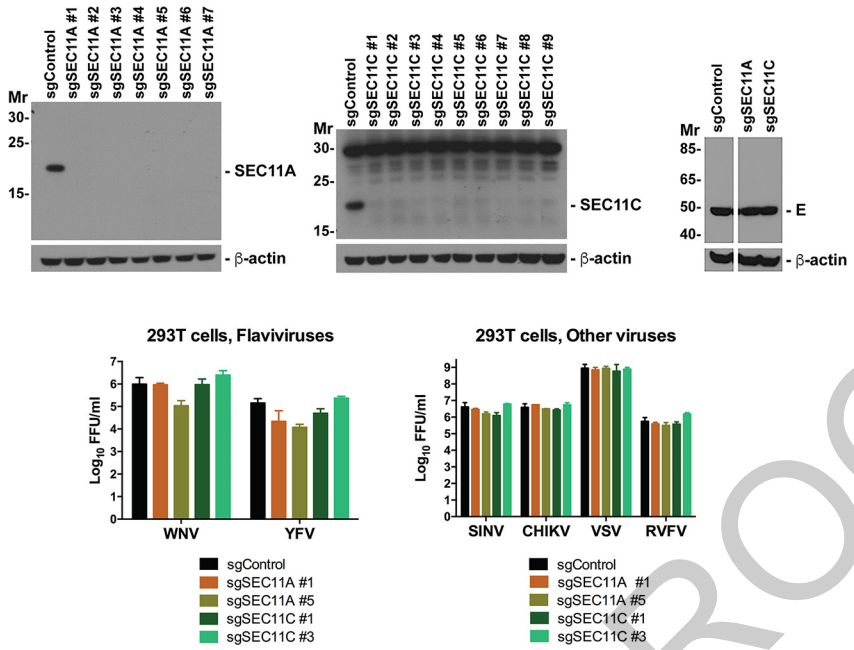
© 2016 Macmillan Publishers Limited. All rights reserved



NOT FINAL PROOF

Extended Data Figure 3 | Viral infection in *SPCS1*$^{-/-}$ cells. **a**–**c**, Alphaviruses replicate and are processed efficiently in 293T cells in the absence of expression of SPCS1. **a**, SINV infection in control and *SPCS1*$^{-/-}$ clonal cells. Cells were infected (MOI 0.01) and supernatants were collected and analysed by FFA. The results are the average of two independent experiments performed in triplicate. **b**, Control and *SPCS1*$^{-/-}$ gene-edited 293T cells were infected with SINV. At the indicated time, lysates were prepared, electrophoresed and western blotted with anti-SINV E2 ascites fluid (ATCC VR-1248AF). **c**, Control or *SPCS1*$^{-/-}$ 293T cells were infected with CHIKV (MOI 5). After 8 h, cells were labelled for 30 min with [^{35}S]cysteine/methionine. Excess cold cysteine/methionine was added for indicated chase times (0, 1 or 4 h). An uninfected control established the specificity of the immunoprecipitation. After ^{35}S labelling, lysates were prepared and immunoprecipitated with anti-E2 monoclonal antibodies (CHK-48). Immunoprecipitates were left untreated (blank) or treated with Endo H (E) or PNGase F (P) for 1 h at 37 °C before SDS–PAGE and fluorography. **d**, Sequencing of *SPCS1* alleles in gene-edited 293T and Huh7 cell clones after puromycin selection and limiting dilution cloning. The sgRNA targeting site and the 'PAM' sequences are highlighted at the top of the wild-type gene, and the sequence of edited alleles are indicated. **e**, Western blotting of bulk-selected or clonal (clone 7) Huh7.5 cells (control and *SPCS1* sgRNA selected) for expression of SPCS1 (~12 kDa). **f**, WNV, HCV, ZIKV, and JEV (Bennett strain) infection in control and SPCS1-deficient Huh7.5 cells. Cells were infected at an MOI of 0.01 (WNV, ZIKV, JEV) or 1 (HCV) and supernatants were collected and analysed by FFA. The results are the average of two independent experiments performed in triplicate. **g**, Control or *SPCS1*$^{-/-}$ Huh7.5 cells were infected at an MOI of 150 for 45 h with a pathogenic JEV isolate (Bennett strain). Lysates were blotted with an anti-JEV E monoclonal antibody. Higher molecular mass bands (E^{hi} and E^{med}) that reacted specifically with the anti-E monoclonal antibody are indicated. One representative experiment of two is shown and loading controls (β-actin) are included. For gel source data, see Supplementary Fig. 1.

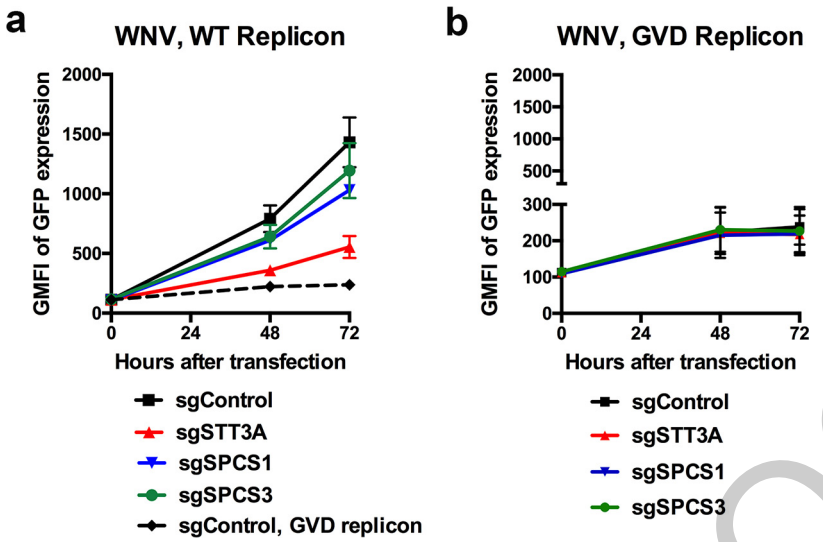
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Gene editing of *SEC11A* and *SEC11C* do not substantively affect infection of several enveloped viruses. Top, 293T cells were treated with the indicated sgRNA and isolated in bulk after puromycin drug selection. Western blotting confirmed gene editing of *SEC11A* (left, 20 kDa) or SEC11C (middle, 22 kDa). No difference in levels or migration pattern of WNV E was observed in *SEC11A* or *SEC11C* gene-edited cells (right) after WNV infection at an MOI of 200 for 24 h. Spaces between the western blots indicate cropping to remove lanes that were not relevant to this experiment. Bottom, control or gene-edited 293T cells were infected with viruses and supernatants were harvested after infection for titration. Left, WNV (MOI 0.01, 72 h) or YFV (MOI 1, 72 h); right, SINV (MOI 0.01, 72 h), CHIKV (MOI 0.01, 36 h), VSV (MOI 0.01, 36 h), or RVFV (MOI 1, 72 h). Results are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

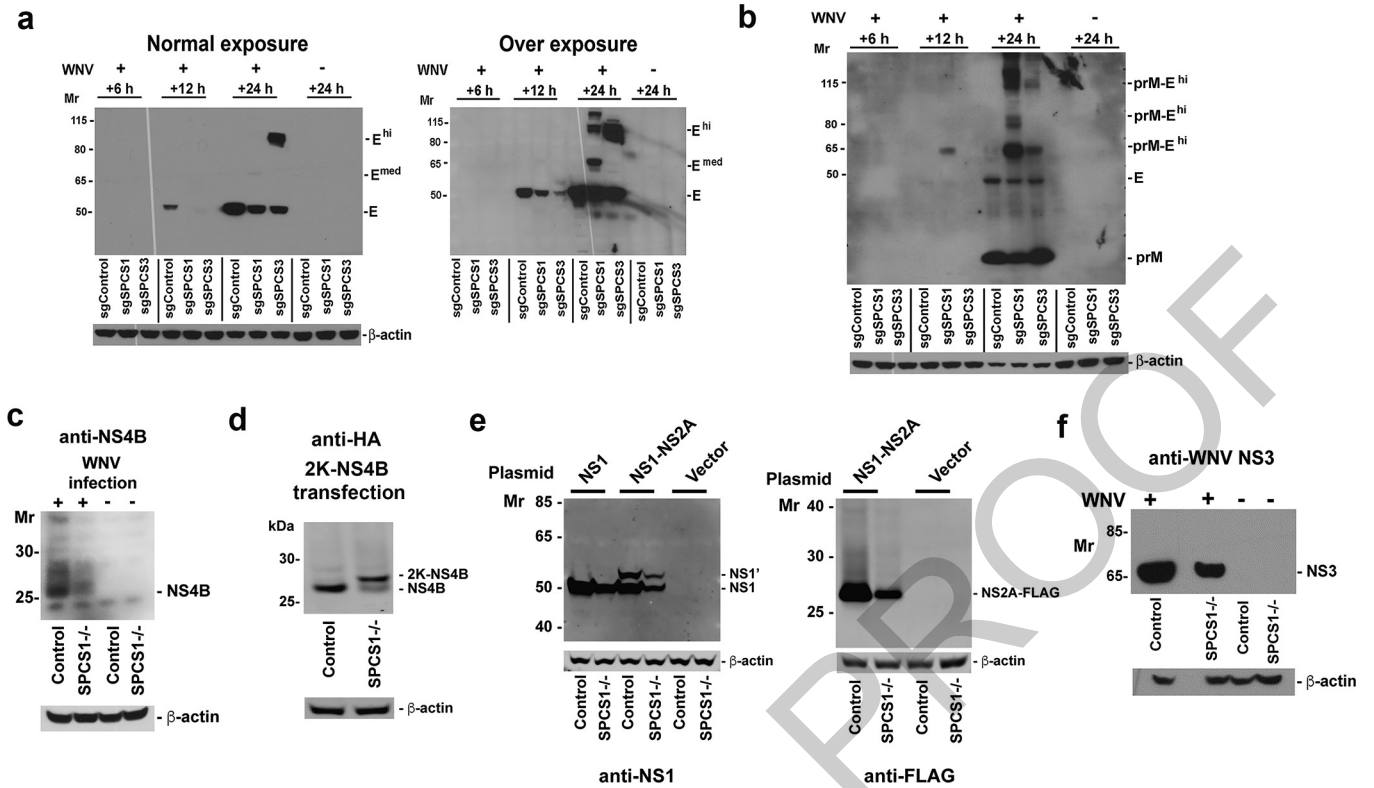
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | Effect of sgRNA on translation and replication of wild-type and NS5 GVD polymerase mutant WNV replicons. A cDNA launched WNV replicon (**a**, wild-type; **b**, GVD polymerase 'dead' mutant) with a minimal CMV promoter (GFP–NS1–NS2A–NS2B–NS3–NS4A–NS4B–NS5) was transfected into the indicated gene-edited 293T cells. At 48 and 72 h after transfection, cells were collected and analysed for GFP expression by flow cytometry. After transfection with the wild-type replicon, WNV replication was lower in *STT3A* gene-edited cells, as determined by ANOVA with a multiple comparisons correction ($P < 0.05$ at 48 and 72 h). Data are the average of three independent experiments. Note, the GVD replicon data with sgRNA control (translation only) are provided for comparison in **a**.

NOT FINAL PROOF

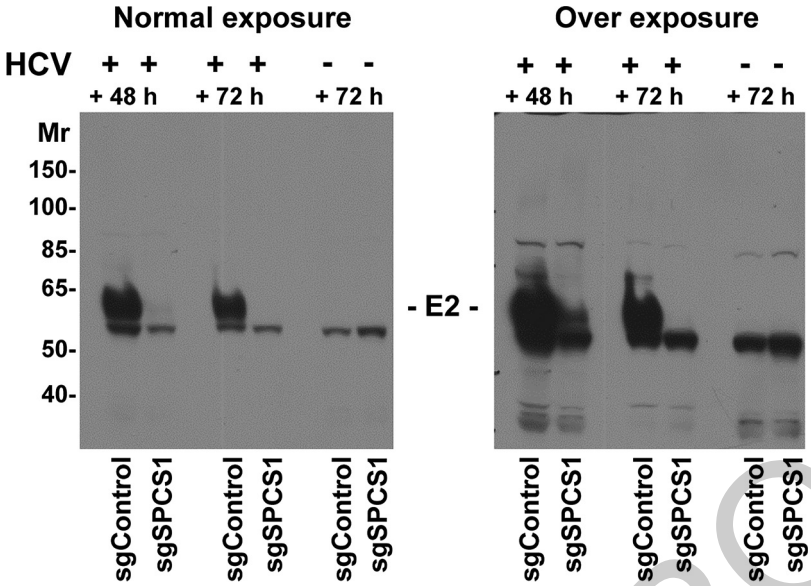
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 6 | Processing of WNV proteins in *SPCS1* and *SPCS3* gene-edited 293T cells. a, Normal (left) and over-exposed (right) western blot in *SPCS1* and *SPCS3* bulk gene-edited 293T cells. The over-exposure is shown to highlight the accumulation of high molecular mass bands that react with anti-E protein antibody. Control, *SPCS1* and *SPCS3* gene-edited 293T cells were infected with WNV or mock-infected for the indicated times. Lysates were western blotted with an anti-E (human E16) monoclonal antibodies. Under these electrophoresis conditions, natively processed E protein migrates at ~50 to 55 kDa. Higher molecular mass bands (E^{med} (probably prM–E) and E^{hi} (probably prM–E–NS1)) that react specifically with the E monoclonal antibody are present only in *SPCS1* and *SPCS3* gene-edited 293T cells. The data are representative of two independent experiments and a loading control (β-actin) is shown. **b**, Western blot of *SPCS1* and *SPCS3* bulk gene-edited 293T cells. Control, *SPCS1*, and *SPCS3* gene-edited 293T cells were infected with WNV or mock-infected for the indicated times. Lysates were western blotted with an anti-prM human monoclonal antibody (CR4293) that recognizes a shared epitope on prM and E. Higher molecular mass bands (prM–E^{hi}) probably represent uncleaved polyprotein forms and are present only in *SPCS1* and *SPCS3* gene-edited 293T cells. The data are representative of two independent experiments and a loading control (β-actin) is shown. **c**, Control or *SPCS1*$^{-/-}$ cells were infected with WNV or left unmanipulated (—) and 24 h later cell lysates were generated and probed with a polyclonal antibody against NS4B. The results are representative of two independent experiments and loading controls (β-actin) are shown. **d**, Control or *SPCS1*$^{-/-}$ cells were transfected with a 2K-NS4B–HA plasmid. One day later, lysates were probed with an anti-HA antibody. Results are representative of two independent experiments and loading controls (β-actin) are shown. Cleaved (NS4B) and uncleaved (2K–NS4B) bands are indicated on the right of the gel. **e**, Control or *SPCS1*$^{-/-}$ cells were transfected with NS1, NS1–NS2A–Flag, or control plasmids. One day later, lysates were probed with anti-NS1 (left) or anti-FLAG (right) antibodies. Cleavage of NS1–NS2A–Flag results in expression of the C-terminal Flag tag exclusively with the residual NS2A protein. The results are representative of three independent experiments and loading controls (β-actin) are shown. Note, expression of NS1–NS2A results in two forms of NS1 (NS1 and NS1′) owing to a ribosomal frameshift event that occurs at a heptanucleotide motif near the beginning of the *NS2A* gene. **f**, Control or *SPCS1*$^{-/-}$ cells were infected with WNV or left uninfected and 24 h later cell lysates were generated and probed with a monoclonal antibody against NS3. The results are representative of two independent experiments and loading controls (β-actin) are shown. For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

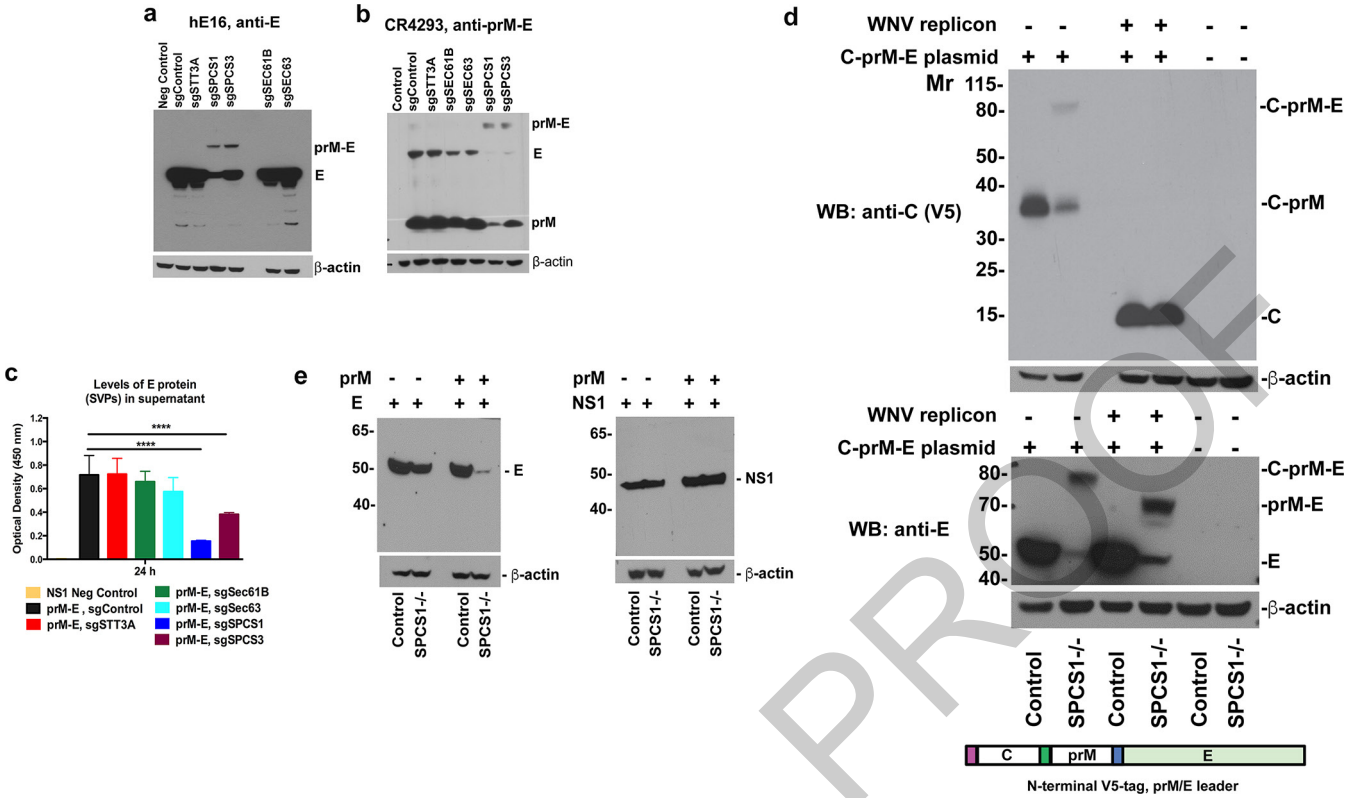
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | Western blotting for HCV E2 in control and *SPCS1* gene-edited Huh7.5 cells. Control or *SPCS1* gene-edited cells were infected with HCV (MOI 5; +) or left untreated (−) and 48 or 72 h later cell lysates were generated and probed with a mouse monoclonal antibody against HCV E2 protein. The results are representative of two independent experiments and a normal and over-exposed blot are shown. For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

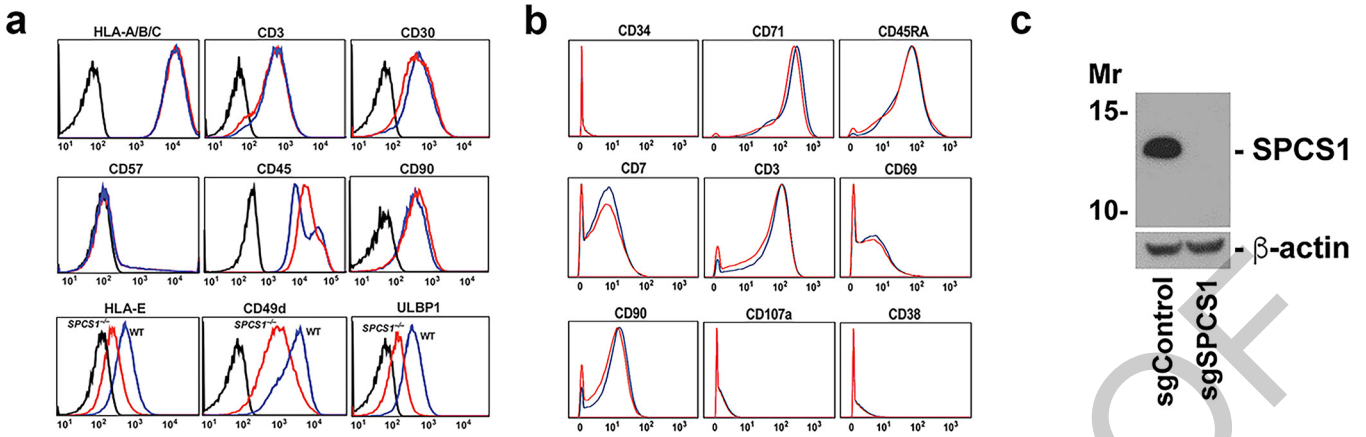
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | Effect of sgRNA on WNV structural protein processing and production. a–**c**, The indicated gene-edited 293T cells were transfected with a plasmid encoding WNV prM–E and subjected to western blotting with hE16 (anti-E) (**a**) or CR4293 (anti-prM-E) (**b**). Note the shift of the prM–E bands to high molecular mass in bulk gene-edited cells with reduced expression of *SPCS1* or *SPCS3*. The results are representative of three independent experiments and a loading control (β-actin) is shown. **c**, 293T cells expressing the indicated sgRNA were transfected with a plasmid encoding prM–E. After 24 h, supernatants were collected and SVPs were quantified by capture ELISA. The results are the average of several independent experiments performed in triplicate. The asterisks indicate SVP levels in the supernatant that are statistically different compared to control cells (****$P < 0.001$, ANOVA with a multiple comparison correction). **d**, Control or *SPCS1*$^{-/-}$ clonal 293T cells were transfected with a single C–prM–E plasmid containing an N-terminal V5 tag fused to C (purple box) and native C–prM (green box) and prM–E (blue box) leader sequences. In some experiments, a cDNA launched WNV replicon was co-transfected to facilitate the cleavage of C from prM by the viral NS2B–NS3 protease. Lysates were prepared 24 h later and probed with an anti-V5 (top) or anti-E (bottom) antibody. Note, two separate gels were run for blotting with ant-V5 (C) and anti-E. One representative experiment of two is shown and a loading control (β-actin) for the top (anti-V5) gel is included. **e**, Control and *SPCS1*$^{-/-}$ 293T cells were transfected with *E* or *NS1* with or without *prM* co-transfection. One day after transfection, cells were collected and lysates were western blotted with antibodies against E (left) or NS1 (right). Molecular mass markers and specific proteins are indicated to the left and right of each gel, respectively. The results are representative of three independent experiments. For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

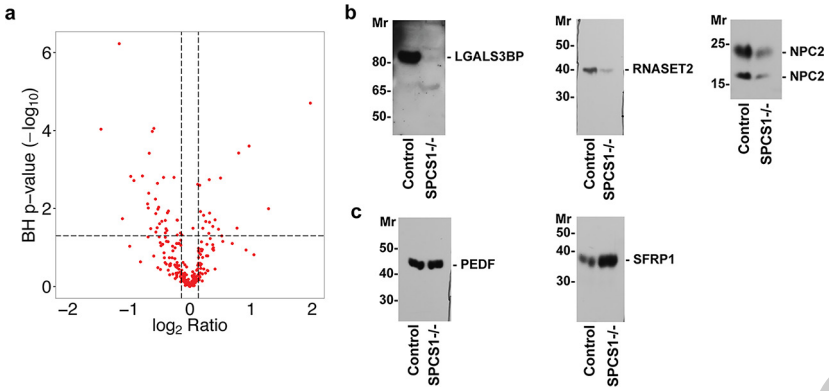
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Expression of immune system antigens on the surface of *SPCS1* gene-edited cells. a, **b**, Control and *SPCS1* gene-edited Jurkat cells were incubated with monoclonal antibodies against the indicated cell surface antigens. After washing, cells were fixed with paraformaldehyde and then processed by flow cytometry (**a**) or mass cytometry (**b**). The histograms are as follows: black, isotype control in wild-type cells; blue, control cells; red, *SPCS1* gene-edited cells. Results are representative of three independent experiments for flow cytometry and one run on a mass cytometer in triplicate. (**c**) Western blotting of bulk-selected Jurkat cells (control and *SPCS1* sgRNA selected) for expression of *SPCS1* (~12 kDa). For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 10 | Secretome analysis in control and *SPCS1*$^{-/-}$ cell supernatants. a, Volcano plot from one-way ANOVA for secreted protein abundances between control and *SPCS1*$^{-/-}$ 293T cells. The areas of dots are proportional to the $\log_2$ standard deviation of protein ratios. The vertical dashed lines delimit fold changes $\pm$ 1.1 and the horizontal dashed line delimits *P* value $\leq$ 0.05. The red dots show secreted proteins using the SP_PIR classification. Values $< \log_2 0$ indicate secreted proteins that show reduced expression in *SPCS1*$^{-/-}$ 293T cells. **b**, **c**, Western blotting of supernatants from control and *SPCS1*$^{-/-}$ 293T cells. **b**, Proteins (LGALS3BP, RNASET2, and NPC2) identified as downregulated in *SPCS1*$^{-/-}$ 293T cells by mass spectrometry (see Supplementary Tables 4 and 5). **c**, Proteins identified as having similar or possibly higher levels in supernatants of *SPCS1*$^{-/-}$ 293T cells. For gel source data, see Supplementary Fig. 1.

NOT FINAL PROOF

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18615

Toremifene interacts with and destabilizes the Ebola virus glycoprotein

Yuguang Zhao[1]*, Jingshan Ren[1]*, Karl Harlos[1], Daniel M. Jones[1], Antra Zeltina[1], Thomas A. Bowden[1], Sergi Padilla-Parra[1,2], Elizabeth E. Fry[1] & David I. Stuart[1,3]

Ebola viruses (EBOVs) are responsible for repeated outbreaks of fatal infections, including the recent deadly epidemic in West Africa. There are currently no approved therapeutic drugs or vaccines for the disease. EBOV has a membrane envelope decorated by trimers of a glycoprotein (GP, cleaved by furin to form GP1 and GP2 subunits), which is solely responsible for host cell attachment, endosomal entry and membrane fusion[1–7]. GP is thus a primary target for the development of antiviral drugs. Here we report the first, to our knowledge, unliganded structure of EBOV GP, and high-resolution complexes of GP with the anticancer drug toremifene and the painkiller ibuprofen. The high-resolution apo structure gives a more complete and accurate picture of the molecule, and allows conformational changes introduced by antibody and receptor binding to be deciphered[8–10]. Unexpectedly, both toremifene and ibuprofen bind in a cavity between the attachment (GP1) and fusion (GP2) subunits at the entrance to a large tunnel that links with equivalent tunnels from the other monomers of the trimer at the three-fold axis. Protein–drug interactions with both GP1 and GP2 are predominately hydrophobic. Residues lining the binding site are highly conserved among filoviruses except Marburg virus (MARV), suggesting that MARV may not bind these drugs. Thermal shift assays show up to a 14 °C decrease in the protein melting temperature after toremifene binding, while ibuprofen has only a marginal effect and is a less potent inhibitor. These results suggest that inhibitor binding destabilizes GP and triggers premature release of GP2, thereby preventing fusion between the viral and endosome membranes. Thus, these complex structures reveal the mechanism of inhibition and may guide the development of more powerful anti-EBOV drugs.

The recent outbreak of EBOV in West Africa, the worst of more than 30 in the past 40 years, comprised more than 28,000 cases and over 11,000 deaths[11]. In the urgent need to find therapeutics, many small compounds and existing Food and Drug Administration (FDA)-approved drugs have been screened *in vitro* or *in silico* (for example, ibuprofen was suggested by docking experiments[12]) to find lead compounds for drug development or repurpose drugs for the treatment of EBOV disease[13–16]. Among these, a set of selective oestrogen receptor modulators (SERMs) stand out as potential inhibitors from *in vitro* and *in vivo* studies[14]; however, their mechanism of action remains largely unknown. Using recombinant EBOV GP we tested whether nine such compounds could directly bind by a thermal-shift assay (Methods). The results show that toremifene in particular markedly decreases the melting temperature (T_m) of EBOV GP, by up to 14 °C at 100 μM (Fig. 1). This contrasts with the action of inhibitors on most protein targets, which tend to increase stability[17], although destabilization has been reported before[18]. Benztropine[19], the G-protein-coupled receptor (GPCR) antagonist, also decreases the T_m of GP by 4 °C, while other compounds, including ibuprofen, showed T_m shifts of less than 2 °C (Fig. 1, Extended Data Fig. 1). The destabilization effects of toremifene and ibuprofen are both pH and concentration dependent (Fig. 1). The binding constants (K_d values) determined by this assay are 16 μM for toremifene and 6 mM for ibuprofen (Extended

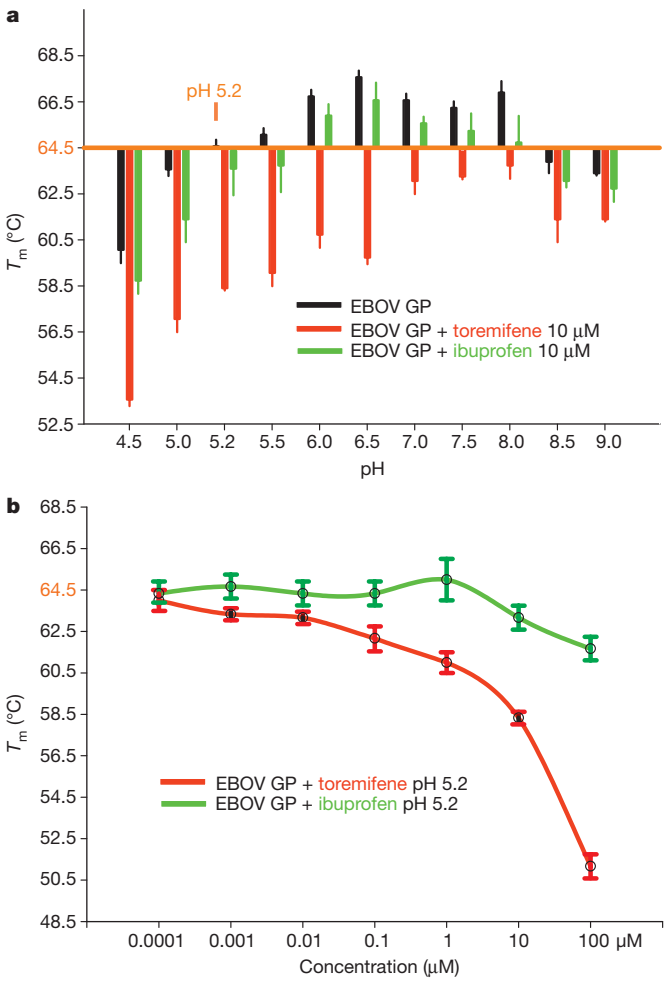


Figure 1 | Summary of thermal-shift assays. a, The effects of toremifene and ibuprofen on the melting temperature of EBOV GP at different pHs. The raw fluorescence traces are shown in Extended Data Fig. 1. The protein melting temperature at pH 5.2 at which the crystals were grown is taken as the reference point. **b**, The melting temperatures of EBOV GP at different concentrations of toremifene or ibuprofen, at pH 5.2. Data are mean ± s.d. ($n = 3$).

[1]Division of Structural Biology, University of Oxford, Wellcome Trust Centre for Human Genetics, Headington, Oxford OX3 7BN, UK. [2]Cellular Imaging Core, University of Oxford, Wellcome Trust Centre for Human Genetics, Headington, Oxford OX3 7BN. [3]Diamond Light Source Ltd, Harwell Science &Innovation Campus, Didcot OX11 0DE, UK.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved

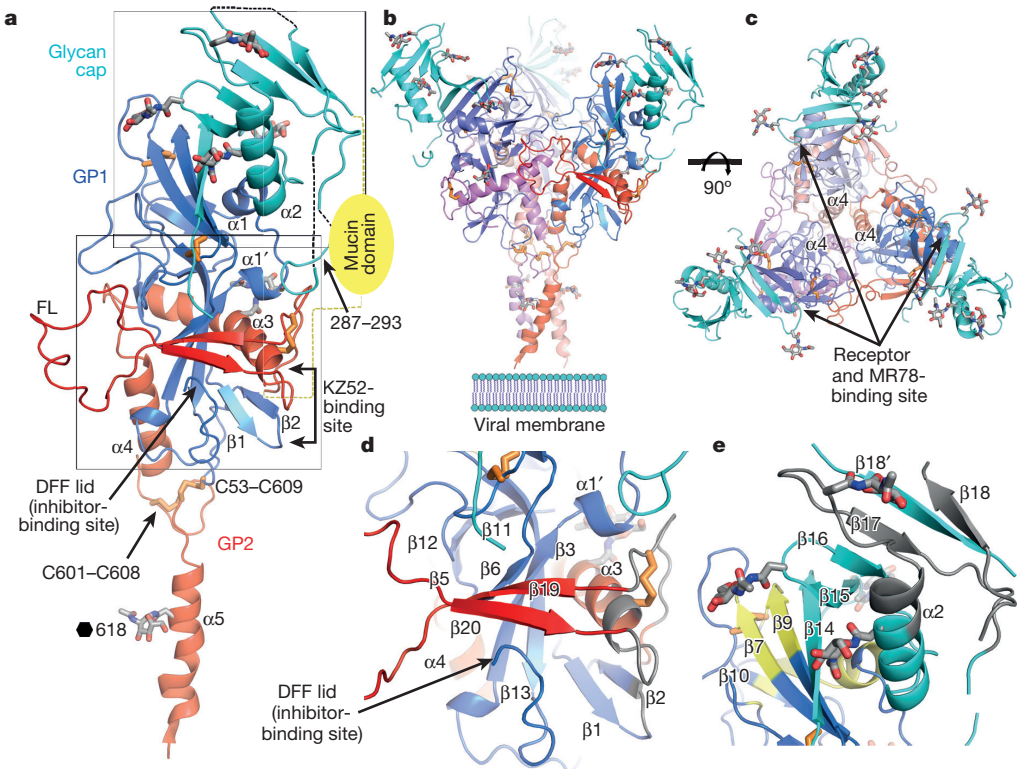


Figure 2 | Overall structure. a, Cartoon diagram of EBOV GP monomer; GP1 is in blue, GP2 is in red, and the glycan cap is in cyan. Secondary structural elements named as described previously[8]. Disulfide bonds shown as orange sticks, glycans in grey. The mucin domain omitted in our construct is shown as a yellow oval. FL, fusion loop. The C-terminal inserted foldon trimerization domain is disordered. **b**, The biological trimer viewed perpendicular to the threefold axis with one monomer coloured as in **a** and the second and third faded for clarity. **c**, The trimer viewed along the threefold, towards the viral membrane. **d**, Close up of the inhibitor binding site. **e**, Close up of the glycan cap and receptor binding site. Areas shown in **d** and **e** are indicated in panel **a**. In **d** and **e** antigenic sites are coloured grey and the receptor binding site yellow.

Data Fig. 1). In a mouse model[14], toremifene appears to be even more potent (half-maximum inhibitory concentration (IC_{50}) ~1 μM).

The crystal structure of unliganded EBOV GP was determined at 2.2 Å resolution, with good R-factors and stereochemistry (Extended Data Table 1). Three copies each of the GP1 and GP2 subunits (Fig. 2a) form the biological trimer around the crystallographic three-fold axis (Fig. 2b, c). This structure, although crystallized at pH 5.2, represents the pre-fusion state of the molecule, with the GP1 receptor-binding site blocked by a glycan cap (Fig. 2e). GP1 is predominantly composed of β-strands, forming a large semi-circular groove at the centre of the subunit that clamps the α3 helix and β19–β20 strands of GP2 (Fig. 2d). The glycan cap is removed in the late endosome by cathepsin B/L to expose the receptor Niemann–Pick disease type C1 (NPC1)-binding site[9,20,21]. GP2 catalyses membrane fusion and contains N-terminal (α3 and α4) and C-terminal (α5) heptad repeats linked by a CX_6CC motif (residues 601–609, Fig. 2a). The C-terminal heptad repeat, disordered in all previously published GP structures[8–10], contributes to the trimer interface in our structure (Fig. 2b, c) and contains N618, which is glycosylated as predicted. The well-ordered CX_6CC motif forms intrasubunit (C601–C608), and intersubunit (C53–C609) disulfide bonds (Fig. 3). Mutation of any of these cysteine residues renders the virus incapable of entering host cells[22]. In the fusion process, GP2 undergoes conformational changes in which α5 refolds onto a helix coalesced from α3 and α4 to form a six-helix bundle[23] (Extended Data Fig. 2). In our pre-fusion structure, the hydrophobic GP2 fusion loop (residues 511–554) (Fig. 2a) projects into a neighbouring monomer and is stabilized in a shallow depression surrounded by loops β4–β5 and β10–β11, and α3. Apart from residues 521–526, which have very loose interactions with the rest of the protein, the fusion loop in this pH 5.2 apo GP is very similar to that in the KZ52 Fab complex crystallized at pH 8.3 (Extended Data Fig. 3). This is in contrast to the large conformational changes reported for the isolated fusion loop at different pHs (ref. 24), suggesting that GP1 maintains GP2 in the pre-fusion state until their separation triggered by receptor binding.

In total, 319 out of 394 Cα atoms in our apo GP structure match with the GP–Fab complex[8], with a root mean squared deviation (r.m.s.d.) value of 1.1 Å (Fig. 3a–d), however, there are marked differences, beyond the C-terminal heptad repeat and CX_6CC motif. The β1–β2 hairpin interacts directly with the KZ52 Fab and is pushed 6 Å inwards in the Fab complex (Fig. 3c). The glycan cap is better ordered in the apo structure, revealing an extra strand, β18′, inserted between β17 and β18, overlapping β18 in the Fab complex but running in the opposite direction (Fig. 3b). Several cross-reactive neutralizing monoclonal antibodies from EBOV survivors bind to the cap[25], suggesting that this conserved epitope is important for antibody-mediated clearance of the virus. Our structure defines this conformational epitope.

In the apo GP structure (excluding the glycan cap), 230 Cα atoms overlay with the GP in the NPC1 receptor complex[9], with an r.m.s.d. of 0.9 Å (Fig. 3e). NPC1 binding draws helix α1 approximately 2 Å towards the receptor, causing the preceding 3_{10}-helix α1′ to unwind, disrupting interactions with α3 of GP2, as described previously[9] (Fig. 3f). These structural changes also break hydrogen bonds from α1′ to the amide groups of residues K510 and N512, disordering the N-terminal residues 502–507 of GP2. The structural differences continue to the other side of α3. In the NPC1-bound structure, the α3 helix starts two residues earlier and the β1–β2 hairpin bends inwards, adopting a conformation similar to that in the KZ52 Fab complex (Fig. 3c, g). In addition, there is a large tunnel between neighbouring monomers (Fig. 4, Extended Data Fig. 4), whose hydrophobic entrance is formed by surrounding residues from the β1–β2 hairpin, β3, β6 and β13 of GP1, and α3, β19 and β20 of GP2. Residues 192–195 (with sequence DFFS, and named DFF lid thereafter) form a tight turn with F193 and F194 plugging the entrance of the tunnel and making tight

© 2016 Macmillan Publishers Limited. All rights reserved

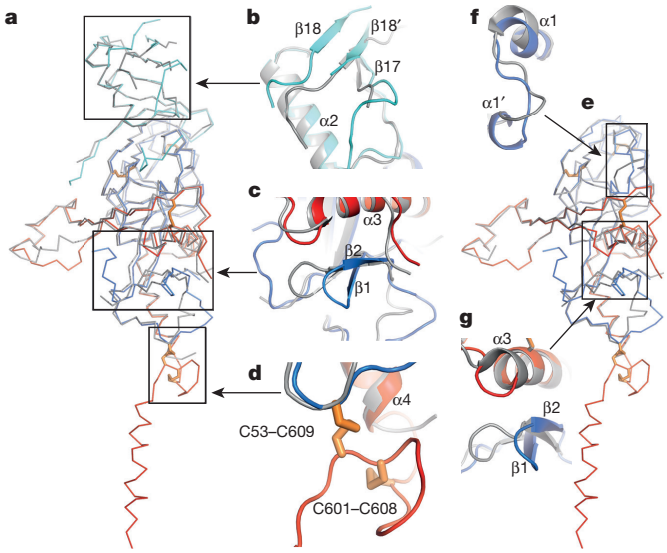


Figure 3 | Structure comparisons. a, Structure of the apo GP monomer compared with the GP of the KZ52 Fab complex. **b**–**d**, Details of the structural differences at the glycan cap (**b**), β1–β2 hairpin (**c**) and the CX_6CC motif (**d**). The apo GP is coloured as in Fig. 1a, the GP in complex with Fab grey. **e**, Comparison of apo GP with GP from the GP–receptor complex shown in same style as **a**. **f**, **g**, close up views of the major structural differences at α1′ and α1 helices (**f**), and β1–β2 hairpin (**g**).

interactions with β19 and β20 (Fig. 4a and Extended Data Fig. 5a). This structure may also hold the putative cleavage site residue 190 in position (Extended Data Fig. 5a)—in the endosome the glycan cap is cleaved to free the receptor-binding site of GP1 (refs 25–29), which also exposes the entrance of the tunnel. Receptor binding is proposed to trigger unwinding of GP2 from GP1 and subsequently lead to membrane fusion[23]. The above structural changes resulting from receptor binding probably all contribute to weaken GP1–GP2 interactions. Both α1′ and α1 are shielded by residues 287–293 and the N563 glycan (which is resistant to enzymatic deglycosylation), perhaps preventing premature release of GP2 (Extended Data Fig. 6).

Structures of GP–toremifene and GP–ibuprofen complexes were obtained by crystal soaking and refined to 2.7 Å resolution (Extended Data Table 1). Both inhibitors have well-defined electron density and bind at the same site at the entrance of the large tunnel by expelling the DFF lid (Fig. 4 and Extended Data Figs 4, 5, 7). In addition to the tunnel entrance residues already mentioned, the tunnel is lined by residues from the N-terminal loop, the β1–β2 hairpin, β2–β3 loop of GP1, and α3 and α4 of GP2 from a neighbouring monomer, and interconnected with the other two tunnels in the trimer at the three-fold axis (Fig. 4b and Extended Data Fig. 4). Y517 makes dominant interactions with toremifene by contacting all three phenyl rings (Fig. 4c). In addition, phenyl ring A of toremifene is fully buried and interacts with V66, L68, L515 and L558, ring B with L186, and ring C with V66 and A101. The ethyl chloride group interacts with L184, L186, M548 and L558, while the dimethylethanamine group points towards the main tunnel and is surrounded by polar/charged residues, including R64, E100, T519, T520 and D522 (Fig. 4c).

Ibuprofen is bound with its isobutyl group partially overlapping the ethyl chloride group of the toremifene but closer to L554. However, its phenyl ring does not overlap any of the rings of toremifene (Extended Data Fig. 5), but makes extensive interactions with M548 (Fig. 4d). The propanoic acid moiety is orientated to make a hydrogen bond to the side chain of R64 and hydrophobic interactions with Y517. Remarkably, ibuprofen was initially suggested to interact with EBOV GP by *in silico* screening, and predicted to dock in a pocket of the mucin domain[12]. A racemic mixture of ibuprofen was used for all experiments, however, we note that the S-isomer (which is also active as a painkiller) binds preferentially.

The flexible region, 521–526, of the fusion loop is stabilized in the two inhibitor-bound structures, but in different conformations compared to apo GP. The most notable conformation changes induced by toremifene are at side chains of M548 and L554, and M548 by ibuprofen (Fig. 4). The residues involved in inhibitor binding are highly conserved across filoviruses, with the exception of MARV (Extended Data Fig. 8), where the DFF lid and its preceding loop are replaced by a helix, and V66 and A101 are substituted by M50 and E85, respectively, partially blocking the binding site[30].

The SERMs tamoxifen, 4-hydroxytamoxifen and clomiphene are less potent inhibitors, despite their chemical similarity to toremifene[14,15]. Compared to the ethyl chloride group of toremifene, the corresponding ethyl group in tamoxifen and chlorine in clomiphene are expected to make weaker interactions with L184, L186, M548 and L558. A partially bound 4-hydroxytamoxifen structure obtained by crystal soaking

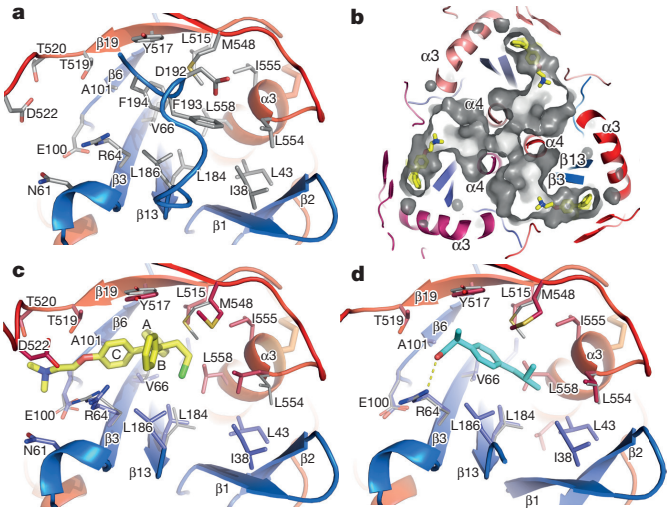


Figure 4 | Inhibitor-binding site. a, Details of the inhibitor-biding site in the apo GP. The backbone is shown as ribbons with GP1 in blue and GP2 in red, side chains as grey sticks. **b**, Tunnels of the GP trimer viewed along the three-fold axis towards the viral membrane. Toremifenes bound at the entrances of the tunnels are shown as yellow sticks. **c**, **d**, Details of protein–inhibitor interactions of the GP–toremifene complex (**c**) and GP–ibuprofen complex (**d**). Toremifene is shown as yellow sticks, ibuprofen as cyan sticks. Protein main chains are shown as ribbons and side chains as sticks (GP1 blue, GP2 red). Side chains in the apo structure with large conformational changes are shown as thinner grey sticks.

© 2016 Macmillan Publishers Limited. All rights reserved

(data not shown) shows the hydroxyl group makes close contacts with G67, shifting the whole inhibitor ~1.0 Å towards the solvent, weakening ring-stacking interactions with Y517 and having no contacts from the ethyl group to L184, L186 and L558 compared to toremifene. Our crystallographic results are in line with the inhibition data[14–16] and our thermal-shift assay (Extended Data Fig. 1e, f). If toremifene and ibuprofen inhibit viral infection by causing premature conversion of GP to the post-fusion conformation or blocking receptor binding, we would expect them to abolish viral fusion. This was confirmed by measuring their effect on the fusion of EBOV GP pseudoviruses as judged by a β-lactamase reporter assay (Extended Data Fig. 9). Benztropine, which decreases the T_m of GP by 4 °C, could not be soaked in our crystals, and needs further investigation (Extended Data Fig. 1f).

Our results pinpoint the binding site of toremifene and ibuprofen on the surface of the GP and reveal that they decrease the stability of the viral GP, and prevent viral fusion. The binding site is different to that predicted for ibuprofen[12], and the information on the binding modes of these compounds and the spare volume at the binding cavity can guide the design of more potent compounds. Finally, our readily grown well-diffracting crystals are suitable for fragment-based screening for different classes of binders for the development of new inhibitors to combat EBOV infection.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 March; accepted 26 May 2016.
Published online 29 June 2016.

1. Takada, A. *et al.* A system for functional analysis of Ebola virus glycoprotein. *Proc. Natl Acad. Sci. USA* **94,** 14764–14769 (1997).
2. Hacke, M. *et al.* Inhibition of Ebola virus glycoprotein-mediated cytotoxicity by targeting its transmembrane domain and cholesterol. *Nat. Commun.* **6,** 7688 (2015).
3. Aleksandrowicz, P. *et al.* Ebola virus enters host cells by macropinocytosis and clathrin-mediated endocytosis. *J. Infect. Dis.* **204** (suppl. 3), S957–S967 (2011).
4. Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477,** 340–343 (2011).
5. Côté, M. *et al.* Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. *Nature* **477,** 344–348 (2011).
6. Nanbo, A. *et al.* Ebolavirus is internalized into host cells *via* macropinocytosis in a viral glycoprotein-dependent manner. *PLoS Pathog.* **6,** e1001121 (2010).
7. Saeed, M. F., Kolokoltsov, A. A., Albrecht, T. & Davey, R. A. Cellular entry of Ebola virus involves uptake by a macropinocytosis-like mechanism and subsequent trafficking through early and late endosomes. *PLoS Pathog.* **6,** e1001110 (2010).
8. Lee, J. E. *et al.* Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454,** 177–182 (2008).
9. Wang, H. *et al.* Ebola viral glycoprotein bound to its endosomal receptor Niemann-Pick C1. *Cell* **164,** 258–268 (2016).
10. Dias, J. M. *et al.* A shared structural solution for neutralizing ebolaviruses. *Nat. Struct. Mol. Biol.* **18,** 1424–1427 (2011).
11. World Health Organization. *Ebola Situation Reports*; http://www.who.int/csr/disease/ebola/situation-reports/en/ (2016).
12. Veljkovic, V. *et al. In silico* analysis suggests repurposing of ibuprofen for prevention and treatment of EBOLA virus disease. *F1000Res.* **4,** 104 (2015).
13. Edwards, M. R. *et al.* High-throughput minigenome system for identifying small-molecule inhibitors of Ebola virus replication. *ACS Infect. Dis.* **1,** 380–387 (2015).
14. Johansen, L. M. *et al.* FDA-approved selective estrogen receptor modulators inhibit Ebola virus infection. *Sci. Transl. Med.* **5,** 190ra79 (2013).
15. Johansen, L. M. *et al.* A screen of approved drugs and molecular probes identifies therapeutics with anti-Ebola virus activity. *Sci. Transl. Med.* **7,** 290ra89 (2015).
16. Kouznetsova, J. *et al.* Identification of 53 compounds that block Ebola virus-like particle entry via a repurposing screen of approved drugs. Emerg. *Microbes Infect.* **3,** e84 (2014).
17. Lea, W. A. & Simeonov, A. Differential scanning fluorometry signatures as indicators of enzyme inhibitor mode of action: case study of glutathione S-transferase. *PLoS One* **7,** e36219 (2012).
18. Cimmperman, P. *et al.* A quantitative model of thermal stabilization and destabilization of proteins by ligands. *Biophys. J.* **95,** 3222–3231 (2008).
19. Cheng, H. *et al.* Inhibition of Ebola and Marburg virus entry by G protein-coupled receptor antagonists. *J. Virol.* **89,** 9932–9938 (2015).
20. Brecher, M. *et al.* Cathepsin cleavage potentiates the Ebola virus glycoprotein to undergo a subsequent fusion-relevant conformational change. *J. Virol.* **86,** 364–372 (2012).
21. Zhao, Y., Ren, J., Harlos, K. & Stuart, D. I. Structure of glycosylated NPC1 luminal domain C reveals insights into NPC2 and Ebola virus interactions. *FEBS Lett.* **590,** 605–612 (2016).
22. Jeffers, S. A., Sanders, D. A. & Sanchez, A. Covalent modifications of the ebola virus glycoprotein. *J. Virol.* **76,** 12463–12472 (2002).
23. Weissenhorn, W., Carfí, A., Lee, K. H., Skehel, J. J. & Wiley, D. C. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. *Mol. Cell* **2,** 605–616 (1998).
24. Gregory, S. M. *et al.* Structure and function of the complete internal fusion loop from Ebolavirus glycoprotein 2. *Proc. Natl Acad. Sci. USA* **108,** 11211–11216 (2011).
25. Flyak, A. I. *et al.* Cross-reactive and potent neutralizing antibody responses in human survivors of natural Ebolavirus infection. *Cell* **164,** 392–405 (2016).
26. Chandran, K., Sullivan, N. J., Felbor, U., Whelan, S. P. & Cunningham, J. M. Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection. *Science* **308,** 1643–1645 (2005).
27. Dube, D. *et al.* The primed ebolavirus glycoprotein (19-kilodalton GP1,2): sequence and residues critical for host cell binding. *J. Virol.* **83,** 2883–2891 (2009).
28. Hood, C. L. *et al.* Biochemical and structural characterization of cathepsin L-processed Ebola virus glycoprotein: implications for viral entry and immunogenicity. *J. Virol.* **84,** 2972–2982 (2010).
29. Schornberg, K. *et al.* Role of endosomal cathepsins in entry mediated by the Ebola virus glycoprotein. *J. Virol.* **80,** 4174–4178 (2006).
30. Hashiguchi, T. *et al.* Structural basis for Marburg virus neutralization by a cross-reactive human antibody. *Cell* **160,** 904–912 (2015).

Acknowledgements We thank Diamond scientists at I02 and I03 for assistance with data collection, T. S. Walter for help with crystallization and thermal-shift essay. Y.Z. was supported by the Biostruct-X project (283570) funded by the EU seventh Framework Programme (FP7), J.R. by the Wellcome Trust, and D.I.S., E.E.F. and K.H. by the UK Medical Research Council (MR/N00065X/1). This is a contribution from the UK Instruct Centre. The Wellcome Trust Centre for Human Genetics is supported by the Wellcome Trust (grant 090532/Z/09/Z). A.Z. is supported by a Marie Curie Fellowship (658363), T.A.B. is supported by the MRC (MR/L009528/1). SP-P is funded by a Nuffield Department of Medicine Leadership Fellowship. D.I.S. is a Jenner Investigator.

Author Contributions Y.Z., J.R. and D.I.S designed the project. Y.Z. made the protein and grew the crystals together with J.R., collected X-ray data and determined the structures. K.H. helped with crystal mounting and data collection. D.M.J. and S.P. carried out cell imaging experiments. A.Z. and T.A.B. provided the cDNA. Y.Z., J.R., E.E.F. and D.I.S. analysed the results and wrote the manuscript in discussions with all authors.

Author Information The atomic coordinates and structure factors have been deposited with the RCSB Protein Data Bank under accession codes 5JQ3 (native GP), 5JQ7 (GP-toremifene) and 5JQB (GP-ibuprofen). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.I.S. (dave@strubi.ox.ac.uk).

Reviewer Information *Nature* thanks E. Saphire, W. Weissenhorn and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Protein cloning, expression and purification. Zaire EBOV (strain Mayinga-76) glycoprotein extracellular domain DNA was synthesized (UniProt entry KB-Q05320). The expression construct GPΔ contains two directly linked sections, amino acids 32–312 and 464–632, with a T42A mutation to eliminate N40 glycosylation. At the N terminus of the protein, the four amino acids ETGR were added from the expression vector pNeosec[31]. At the C terminus, a foldon trimerization sequence from the bacteriophage T4 fibritin and a 6× His tag were added with the sequence: GSGYIPEAPRDGQAYVRKDGEWVLLSTFLGTHHHHHH. The endotoxin free pNeosec-GPΔ plasmid was transiently transfected into the human embryonic kidney HEK293T (ATCC CRL11268) cells with polyethylenimine (PEI, molecular mass 25 kDa, Sigma). To inhibit the formation of complex glycosylation, the mannosidase inhibitor kifunensine (Cayman Chemical) was added to a final concentration of 5 μM. After 5 days of transfection, the conditioned media was collected, dialysed against PBS and incubated with talon beads (Takara Bio Europe SAS) at 15 °C for 1 h with gentle shaking. The beads were collected and washed with PBS plus 5–10 mM imidazole. The protein was eluted with 200 mM imidazole in PBS and further purified by size exclusion chromatography with a Superdex 200 HiLoad 16/600 column (GE Healthcare) and a buffer of 10 mM MES, pH 5.2, 150 mM NaCl.

Thermal-shift assay. 10 μl of solution containing 2 μM glycosylated EBOV GP protein, buffered by the addition of 10 μl of 850 mM sodium malonate at the desired pH, was mixed with 10 μl of 15× SYPRO Orange dye (Thermo Fisher Scientific), along with 10 μl of 50 μM compound in 10% DMSO or just 10% DMSO. The mixture was made up to a total volume of 50 μl. Samples were placed in a semi skirted 96-well PCR plate (4 Titude), sealed and heated in an Mx3005p qPCR machine (Stratagene, Agilent Technologies) from 24.5 to 98.5 °C at a rate of 1 °C min^{-1}. Fluorescence changes were monitored with excitation and emission wavelengths at 492 and 610 nm, respectively. Reference wells, that is, solutions without chemical, but with the same amount of DMSO, were used to compare the melting temperature (T_m). Experiments were carried out in triplicate. Compounds tested included toremifene, tamoxifen, 4-hydroxyltamoxifen, anastrozole, benztropine, clomipramine, ibuprofen, diacylglycerol kinase inhibitor and benzodiazepine. The SERMs endoxifen and clomiphene could not be tested since they bind to SYPRO Orange directly.

Ebola pseudovirus production and titration. HIV-1-derived pseudoviruses expressing the Ebola virus envelope glycoproteins (EBOV pseudoparticle, EBOVpp) were produced as described previously[32]. HEK-293T cells in T175 flasks were transfected with 2 μg pR8ΔEnv, 2 μg BlaM-Vpr, 1 μg pcREV and 3 μg of plexm-EBOV_GP plasmids (containing Zaire EBOV GP residues 1–676 under control of β-actin/CMV chimaeric promoter). After 8 h of transfection, the medium was replaced by fresh DMEM with 10% FBS. Viral supernatants were collected and concentrated using the Lenti-X Concentrator (Clontech). Virus titres were determined by infecting TZM-bl cells (PTA-5659, no mycoplasma contamination) with a serial dilution of concentrated pseudovirus followed by a β-Gal assay. Since the TZM-bl cells contain a β-Gal expression cassette with an HIV-1-induced promoter infected cells can be identified through hydrolysis of X-gal[33].

BlaM assay and analysis. The β-lactamase assay[33] was applied to assess EBOVpp fusion. 24 h before the assay TZM-bl cells were plated at 4×10^4 cells per well in black clear-bottomed 96-well plates. On the day of assay, cells were cooled on ice before the addition of EBOVpp (MOI 0.5), then centrifuged at 2,100*g* for 30 min at 4 °C. Viral supernatants were removed and cells washed with PBS. Then, 100 μl of DMEM plus 10% FBS containing toremifene (15 μM or 1.5 μM), ibuprofen (150 μM or 15 μM), or the same amount of solvent for the above compounds (5% DMSO) was added to each well before placing in a 37 °C, CO_2 incubator to initiate viral entry. After 120 min, cells were loaded with 1× CCF2-AM from the LiveBLAzer FRET—B/G Loading Kit (Life Technologies) and incubated at room temperature in the dark for 2 h. After CCF2-AM removal, cells were washed with PBS and fixed with 2% PFA before viewing. Cells were excited using a 405 nm continuous laser (Leica) and the emission spectra between 430 and 560 nm were recorded pixel by pixel (512 × 512) using a Leica SP8 X-SMD microscope with a 20× objective. The ratio of blue emission (440–480 nm, cleaved CCF2-AM) to green (500–540 nm, uncleaved CCF2-AM) was then calculated pixel by pixel using a customized macro[34] for ImageJ (http://imagej.nih.gov/ij/) with ten different observation fields for each condition. A blue/green threshold (fusion threshold) was set using HIV_{NoEnv} virions containing Vpr-BlaM as a background control to provide a fusion detection limit that corresponded to 0.75 ± 0.05 BlaM ratio. The fusion threshold was calculated recovering the signal (blue/green intensity ratio) coming from individual cells plus 2× s.d. from ~200 cells in the observation field. This threshold was then applied to all conditions. Cells above the threshold were pseudocoloured in red and cells below the threshold pseudocoloured in blue. 'Red' cells were then compared with blue cells (non-fusogenic) as an accurate measure of fusion in different conditions.

Crystallization and inhibitor soaking. For crystallization, the GP was treated with endo-β-acetylglucosaminidase F1 at 37 °C for 1 h, further purified with size-exclusion chromatography and concentrated to 10–12 mg ml^{-1}. Crystallization screen experiments were carried out using the nanolitre sitting-drop vapour diffusion method in 96-well plates as previously described[35]. Crystals were initially obtained from Hampton Research PEGRx 1 screen, condition 37 containing 10% (w/v) PEG 6,000 and 0.1 M sodium citrate tribasic dihydrate (pH 5.0). The best crystals were grown in the optimized condition containing 9% (w/v) PEG 6,000 and 0.1 M sodium citrate tribasic dihydrate at pH 5.2.

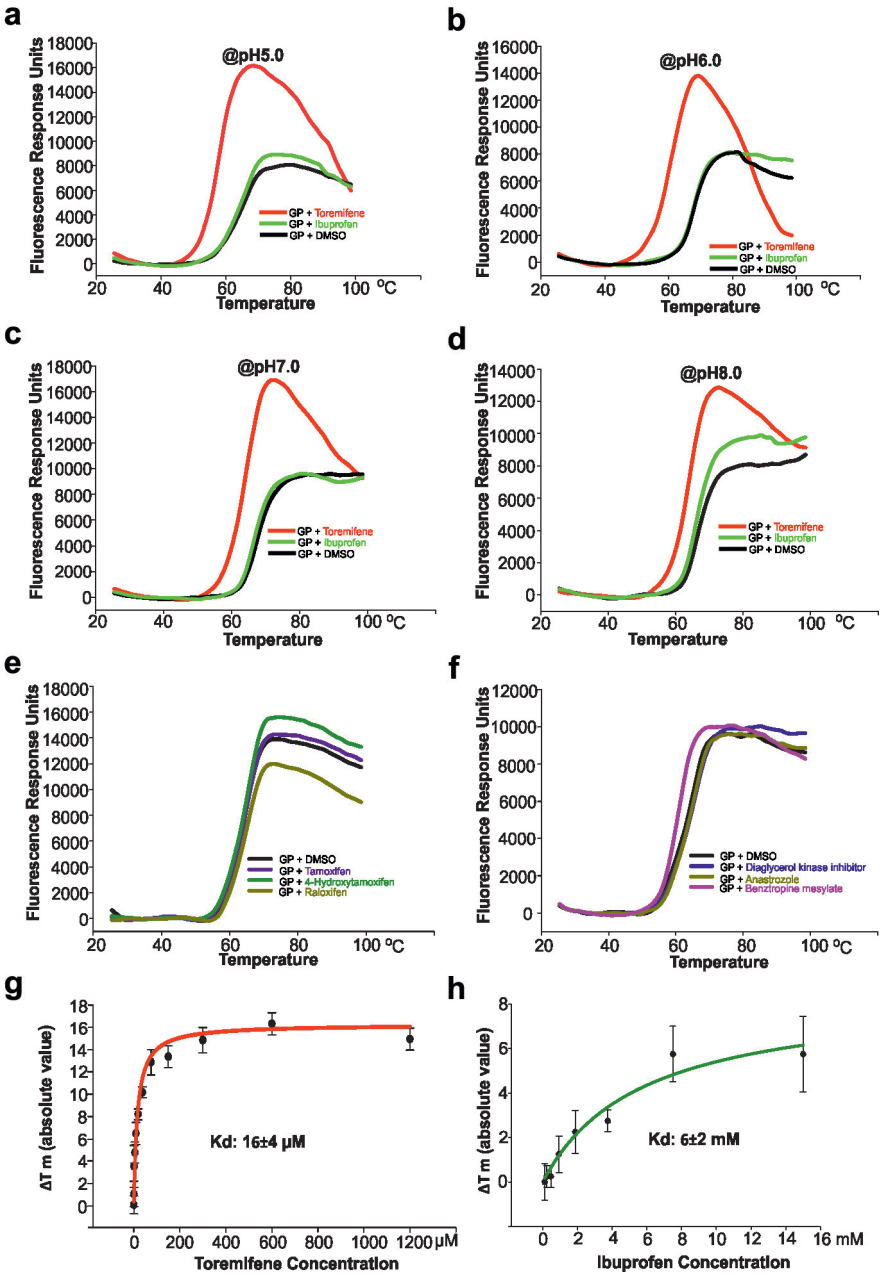
To obtain GP–inhibitor complexes, crystal-soaking experiments were performed. Crystal-soaking solutions were prepared by first dissolving the inhibitors in 100% DMSO and then diluting with 15% (w/v) PEG 6,000 and 0.1 M sodium citrate tribasic dihydrate (pH 5.0) to a final DMSO concentration of 10%. The inhibitor concentration was typically from 1 to 10 mM depending on solubility. Crystals were soaked in the above solutions for between 20 min and 5 h. Crystals soaked for longer normally diffracted less well and the crystals from which the GP–toremifene and GP–ibuprofen complex structures were obtained were only soaked for 20 min.

Data collection and structure determination. Crystals were cryo-protected using solutions containing 75% crystallization liquor (or inhibitor soaking solution) and 25% (v/v) glycerol and frozen in liquid nitrogen before data collection. All data were collected from frozen crystals at 100 K. Data were acquired as 0.1° images on PILATUS 6M detectors at Diamond Light Source, using beamline I03 for native data (exposure time 0.1 s per frame, beam size 80 × 20 μm and 30% beam transmission), and I02 for inhibitor soaked crystals (exposure time 0.05 s per frame, beam size 90 × 25 μm and 40% beam transmission). Diffraction images were indexed, integrated and scaled with the automated data processing program Xia2-3dii[36]. The native data set was collected from four crystals to 2.23 Å resolution with 58-fold redundancy. A total of 7 inhibitors were soaked, including toremifene, tamoxifen, 4-hydroxyltamoxifen, raloxifene, clomiphene, ibuprofen and benztropine, and diffraction data were collected with resolutions ranging from 3.5 to 2.3 Å.

The crystals belong to space group *R*32 with unit cell dimensions $a = b = 114.0$ Å and $c = 307.0$ Å approximately. The apo structure was determined by molecular replacement with MOLREP[37] using the GP structure of the GP-KZ52 Fab complex (PDB ID, 3CSY) as a search model. There is one GP molecule in the crystal asymmetric unit. The biological trimer is formed by a crystallographic three-fold axis. Structure refinement used REFMAC[38] and models were rebuilt with COOT[39]. The apo structure was refined to 2.23 Å resolution with an R_{work} value of 0.223 ($R_{free} = 0.251$) and good stereochemistry. Close examination of the data from inhibitor-soaked crystals showed that only toremifene and ibuprofen were fully bound with GP, and structures were refined to resolutions of 2.69 Å and 2.68 Å, respectively. 4-hydroxyltamoxifen was only bound with partial occupancy (data not shown). Data collection and structure refinement statistics are given in Extended Data Table 1. Structural comparisons used SHP[40], figures were prepared with PyMOL[41].

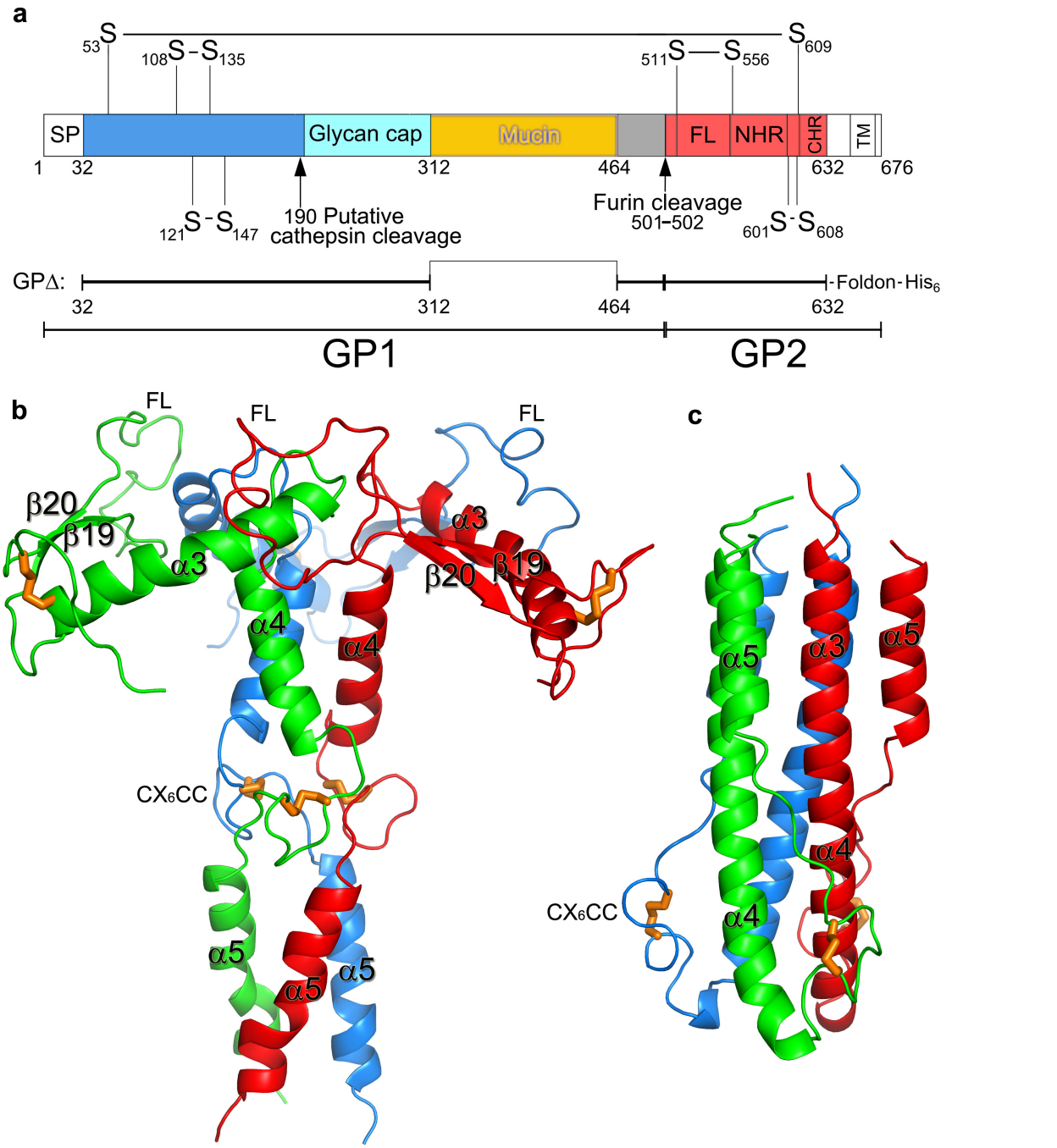
31. Zhao, Y., Ren, J., Padilla-Parra, S., Fry, E. E. & Stuart, D. I. Lysosome sorting of β-glucocerebrosidase by LIMP-2 is targeted by the mannose 6-phosphate receptor. *Nat. Commun.* **5,** 4321 (2014).
32. Desai, T. M. *et al.* IFITM3 restricts influenza A virus entry by blocking the formation of fusion pores following virus-endosome hemifusion. *PLoS Pathog.* **10,** e1004048 (2014).
33. Cavrois, M., De Noronha, C. & Greene, W. C. A sensitive and specific enzyme-based assay detecting HIV-1 virion fusion in primary T lymphocytes. *Nat. Biotechnol.* **20,** 1151–1154 (2002).
34. Jones, D. M. & Padilla-Parra, S. Imaging real-time HIV-1 virion fusion with FRET-based biosensors. *Sci. Rep.* **5,** 13449 (2015).
35. Walter, T. S. *et al.* A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr. D* **61,** 651–657 (2005).
36. Winter, G., Lobley, C. M. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D* **69,** 1260–1273 (2013).
37. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66,** 22–25 (2010).
38. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374,** 300–321 (2003).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
40. Stuart, D. I., Levine, M., Muirhead, H. & Stammers, D. K. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *J. Mol. Biol.* **134,** 109–142 (1979).
41. DeLano, W. L. & Lam, J. W. PyMOL: A communications tool for computational models. *Abstr. Pap. Am. Chem. Soc.* **230,** U1371–U1372 (2005).

© 2016 Macmillan Publishers Limited. All rights reserved



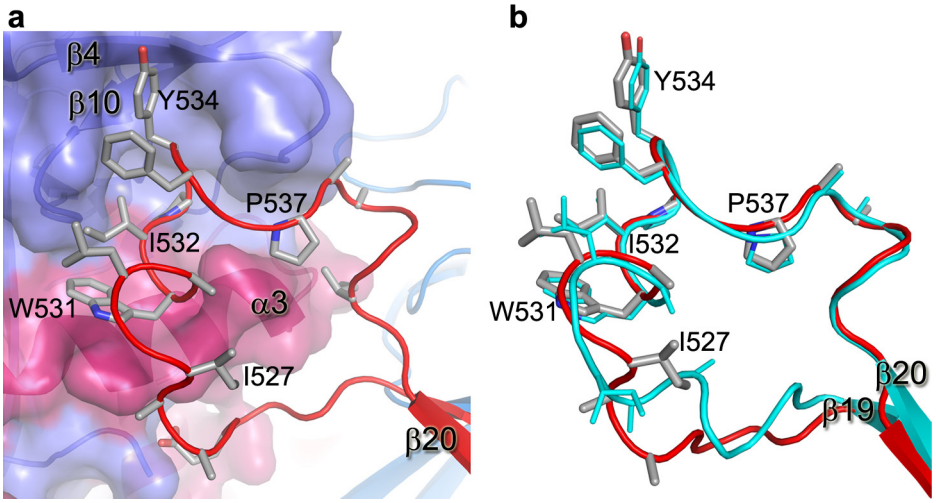
Extended Data Figure 1 | Thermal-shift essay. Representative thermal melt curves of EBOV GP with 10 μM compounds and 2% DMSO. **a**–**d**, Melting curves of EBOV GP with toremifene, ibuprofen or protein alone at pH 5.0, 6.0, 7.0 and 8.0, respectively. **e**, Small effects of SERM inhibitors tamoxifen, 4-hydroxytamoxifen and raloxifen on the melting temperature of EBOV GP shown at pH 5.2. **f**, Melt curves of EBOV GP with diaglycerol kinase inhibitor, anastrozole and benztropine mesylate at pH 5.2. **g**, **h**, Shifts in melting temperature (ΔT_m °C in absolute value) were plotted against different concentrations of toremifene (**g**) or ibuprofen (**h**) at pH 5.2. Data are mean ± s.d. ($n = 4$). The affinity constant K_d is calculated by a ligand binding 1:1 saturation fitting with the SigmaPlot version 13 (Systat Software Inc.).

© 2016 Macmillan Publishers Limited. All rights reserved



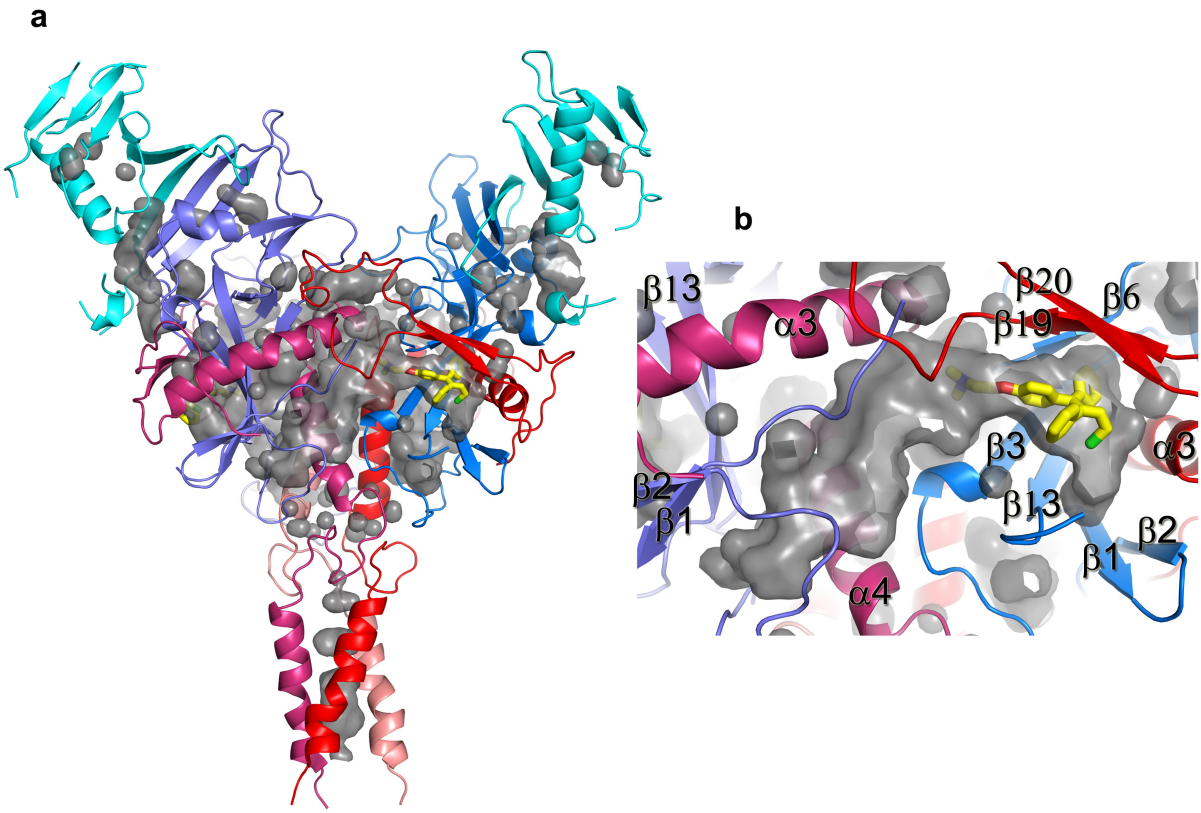
Extended Data Figure 2 | Structural organization of EBOV GP and GP2 structure. a, Scheme showing the structural organization of EBOV GP. FL, fusion loop; NHR and CHR, N- and C-terminal heptad repeats; SP, signal peptide; TM, transmembrane helix. The GPΔ construct used for structure determination is made by deleting residues 313–463 of the GP mucin domain and residues 633–676. Residue 312 is directly linked to 464. A foldon trimerization peptide and a 6× His tag are added at the C terminus. **b**, The GP2 trimer in the prefusion state (current structure). The trimer is shown as cartoon representation with the monomers coloured in red, green and blue, respectively. Disulfide bonds are shown as orange sticks. **c**, The six-helix bundle of GP2 in the post-fusion state.

© 2016 Macmillan Publishers Limited. All rights reserved



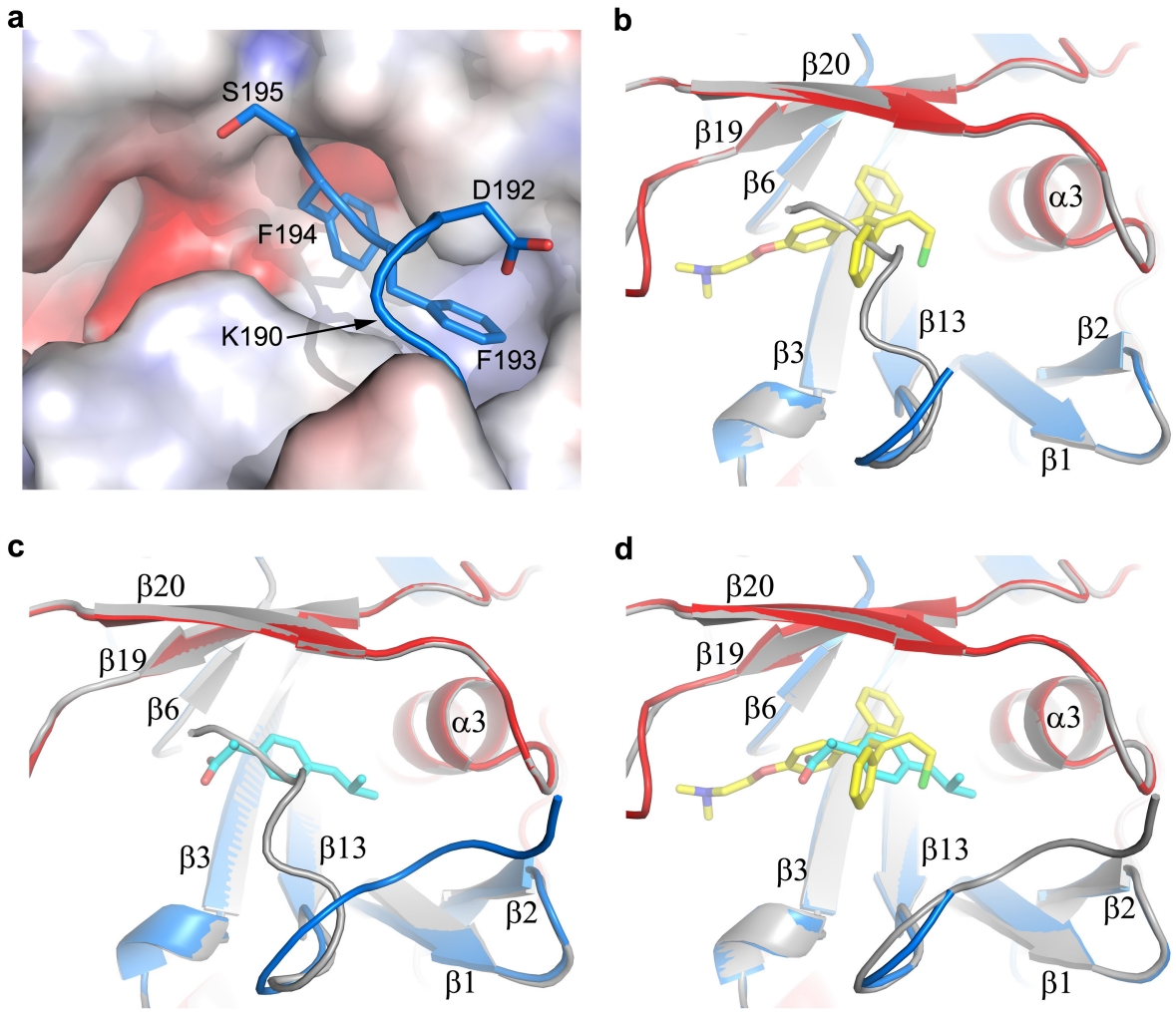
Extended Data Figure 3 | The fusion loop. a, The fusion loop that connects β19 and β20 of GP2 projects onto a shallow depression on the surface of a neighbouring monomer. The fusion loop is shown as a red coil with side chains drawn as grey sticks, the neighbouring monomer is shown in semi-transparent surface representation. **b**, Comparison of the fusion loop in the apo GP (red and grey) obtained at pH 5.2 with that in the KZ52 Fab complex (cyan) obtained at pH 8.3.

© 2016 Macmillan Publishers Limited. All rights reserved



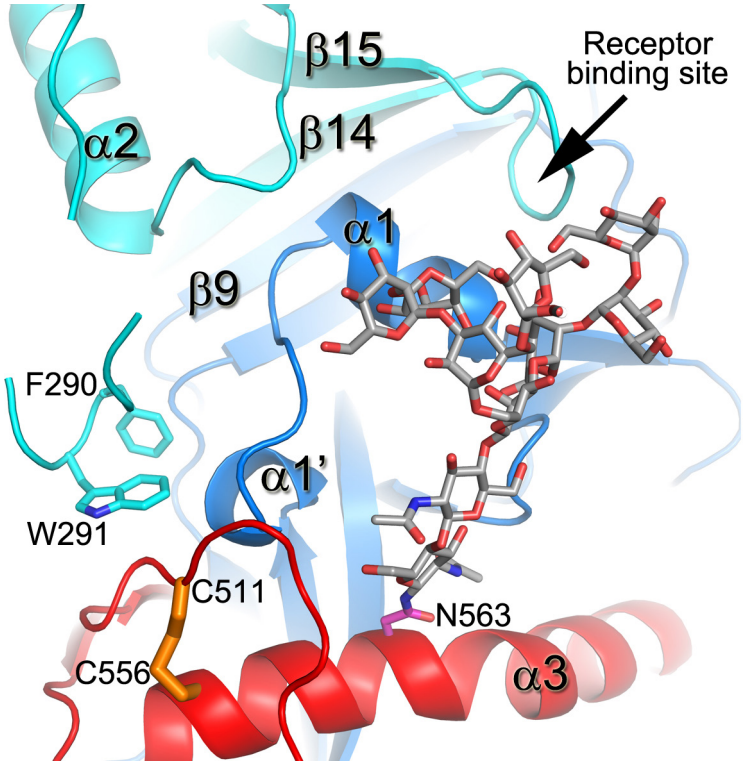
Extended Data Figure 4 | Pockets and tunnels in EBOV GP trimer. **a**, The several small pockets and three large tunnels in the GP trimer shown as grey surfaces. Protein backbones are drawn as ribbons and coloured as in Fig. 2 of the main text. A toremifene is bound at the entrance of each large tunnel and shown as yellow sticks. **b**, Close up view of a tunnel. Each tunnel is bordered by secondary structure elements from two neighbouring monomers.

© 2016 Macmillan Publishers Limited. All rights reserved



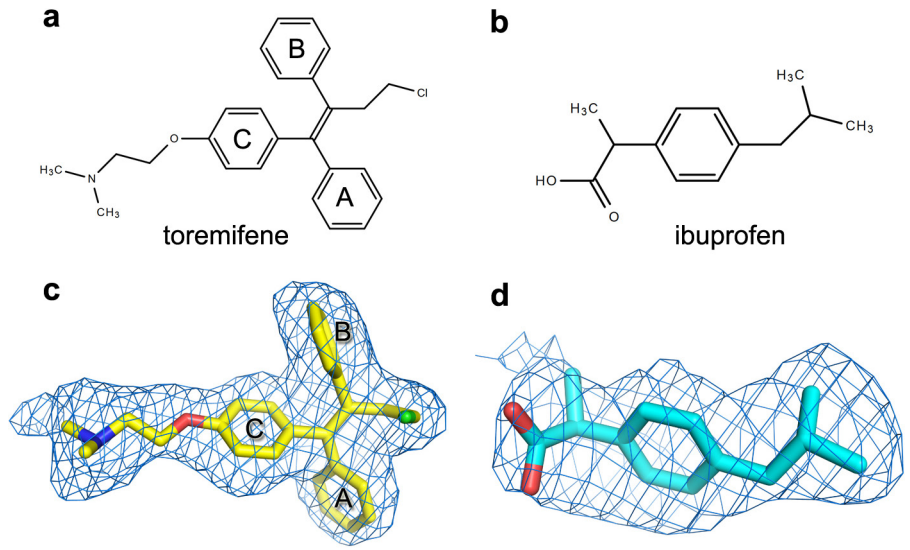
Extended Data Figure 5 | The inhibitor-binding site. a, The DFF lid (residues 192–194, blue coil for main chain and sticks for side chains) nestles at the entrance of the large tunnel in the apo structure. The rest of the protein is shown as an electrostatic surface. The putative cathepsin cleavage site at residue 190 is indicated by an arrow. **b**, **c**, Toremifene (yellow sticks in **b**) and ibuprofen (cyan sticks in **c**) bind at the same site by expelling the DFF lid. In both panels, the inhibitor bound structure is shown in blue (GP1) and red (GP2), the apo GP in grey. **d**, Comparing the binding modes of toremifene and ibuprofen. The toremifene-bound structure is shown in blue and red, the ibuprofen bound structure in grey.

© 2016 Macmillan Publishers Limited. All rights reserved



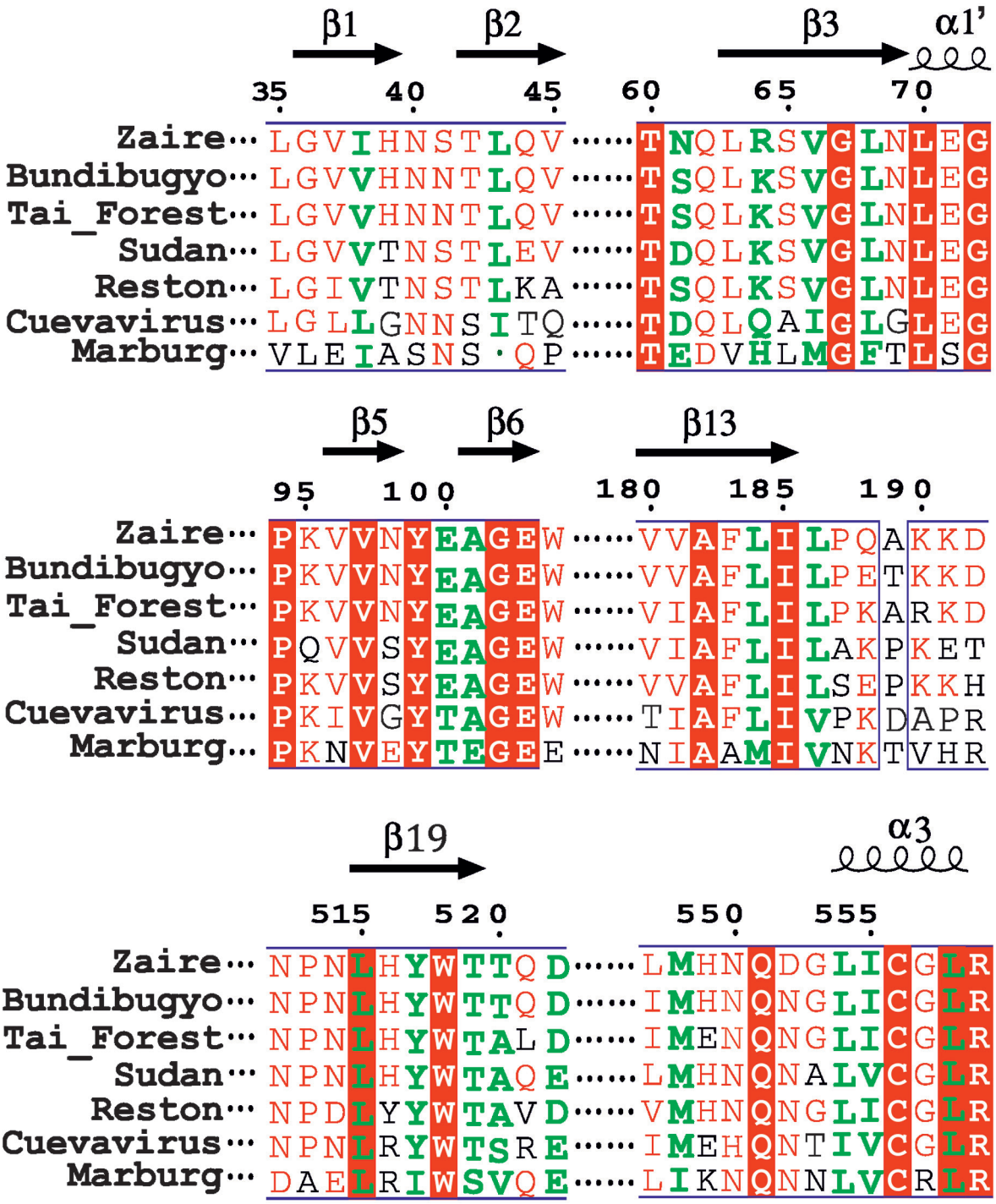
Extended Data Figure 6 | The environment of α1′ and α1 helices. The surfaces of α1′ and α1 helices, which undergo large conformational changes upon receptor binding, are protected by the 287–293 loop from the glycan cap domain and the N563 glycan from GP2 in the apo GP. The glycan is modelled as Man9GlcNAc2.

© 2016 Macmillan Publishers Limited. All rights reserved



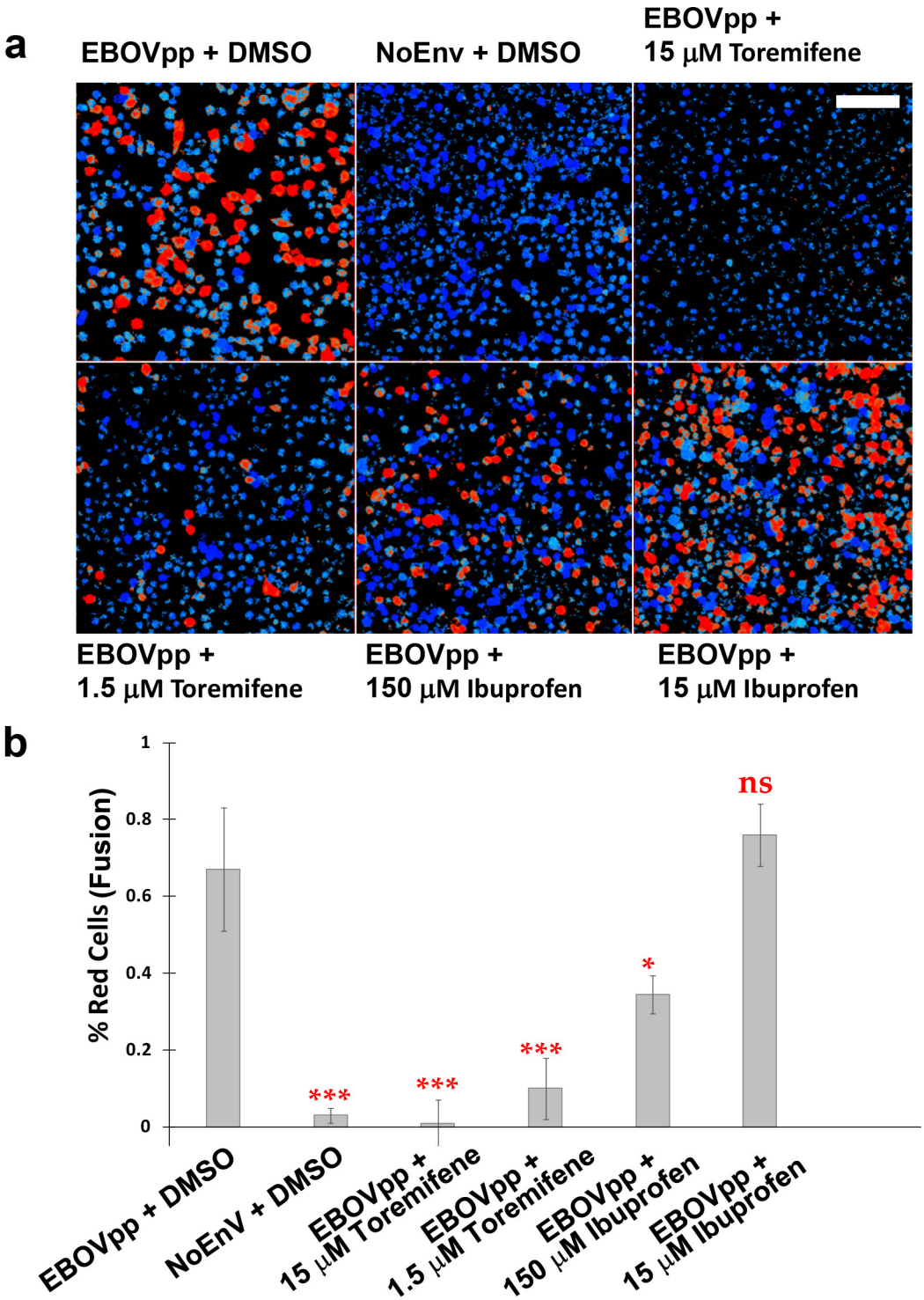
Extended Data Figure 7 | Chemical structures and electron density maps. a, **b**, The chemical structures of toremifene (**a**) and ibuprofen (**b**). **c**, **d**, $|F_o - F_c|$ omit electron density maps for toremifene (**c**) and ibuprofen (**d**) contoured at 3σ.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | Sequence alignment of filovirus GPs. Amino acid sequence alignment of 7 filovirus GPs around the inhibitor-binding site. The amino acids that form contacts with toremifene or ibuprofen are coloured in green. Numbering corresponds to the full length Zaire EBOV GP, conserved residues are shown in a red background. Secondary structure elements are labelled on the top.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Toremifene and ibuprofen inhibit fusion of Ebola GP pseudovirus particles. a, CCF2-loaded TZM-bl cells were exposed to EBOV pseudoparticles (EBOVpp) or control particles lacking envelope proteins (NoENV) at 4 °C to synchronise binding and receptor engagement before fusion was initiated by shifting cells to 37 °C in the presence of toremefine (15 μM and 1.5 μM), ibuprofen (150 μM and 15 μM), or just the solvent (5% DMSO). After 2 h incubation, cells were loaded with the CCF2-AM FRET biosensor, fixed and the ratio of blue (440–480 nm, cleaved CCF2-AM) to green (500–540 nm, uncleaved CCF2-AM) fluorescence measured. Cells are pseudocoloured according to this ratio: blue represents no fusion, red represents fusion. Scale bar: 80 μm. **b**, The percentage of fusogenic cells (red versus blue) was calculated taking the average max value coming from the negative control as a threshold for fusion, data are means ± s.d. ($n = 10$). *$P \leq 0.05$, ***$P \leq 0.001$ (unpaired t-test, compared to the EBOV plus DMSO control). ns, not significant ($P > 0.05$). Error bars represent s.d.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Data collection and refinement statistics

	Native GP	GP-toremifene	GP-ibuprofen
Data collection			
Space group		*R*32	
Cell dimensions			
a, *b*, *c* (Å)	114.3, 114.3, 307.4	113.5, 113.5, 306.9	113.8, 113.8, 306.2
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	94.2–2.23 (2.29–2.23)*	51.2–2.69 (2.76–2.69)	82.8–2.68 (2.75–2.68)
R_{merge}	0.204(---)	0.079 (---)	0.143 (---)
$I / \sigma I$	17.4(1.3)	20.0 (1.9)	14.7(1.5)
Completeness (%)	100(100)	99.9 (100)	99.9(100)
Redundancy	57.8(15.4)	9.8 (8.6)	9.8(8.3)
Refinement			
Resolution (Å)	94.2–2.23	51.2–2.68	82.8–2.68
No. reflections	36035/1865	20449/1090	20734/1107
R_{work} / R_{free}	0.226/0.241	0.201/0.245	0.199/0.235
No. atoms			
Protein	3129	3034	3025
Ligand/glycan/ion	143	151	140
Water	119	53	73
B-factors			
Protein	39	62	55
Ligand/glycan/ion	99	115	96
Water	54	70	55
R.m.s. deviations			
Bond lengths (Å)	0.007	0.007	0.007
Bond angles (°)	1.3	1.3	1.3

*Values in parentheses are for highest-resolution shell.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18317

A core viral protein binds host nucleosomes to sequester immune danger signals

Daphne C. Avgousti[1,2], Christin Herrmann[2,3], Katarzyna Kulej[1,2], Neha J. Pancholi[2,3], Nikolina Sekulic[4,5]†, Joana Petrescu[2,6], Rosalynn C. Molden[5], Daniel Blumenthal[7], Andrew J. Paris[8], Emigdio D. Reyes[1,2], Philomena Ostapchuk[9], Patrick Hearing[9], Steven H. Seeholzer[10], G. Scott Worthen[11], Ben E. Black[4,5], Benjamin A. Garcia[4,5] & Matthew D. Weitzman[1,2]

Viral proteins mimic host protein structure and function to redirect cellular processes and subvert innate defenses[1]. Small basic proteins compact and regulate both viral and cellular DNA genomes. Nucleosomes are the repeating units of cellular chromatin and play an important part in innate immune responses[2]. Viral-encoded core basic proteins compact viral genomes, but their impact on host chromatin structure and function remains unexplored. Adenoviruses encode a highly basic protein called protein VII that resembles cellular histones[3]. Although protein VII binds viral DNA and is incorporated with viral genomes into virus particles[4,5], it is unknown whether protein VII affects cellular chromatin. Here we show that protein VII alters cellular chromatin, leading us to hypothesize that this has an impact on antiviral responses during adenovirus infection in human cells. We find that protein VII forms complexes with nucleosomes and limits DNA accessibility. We identified post-translational modifications on protein VII that are responsible for chromatin localization. Furthermore, proteomic analysis demonstrated that protein VII is sufficient to alter the protein composition of host chromatin. We found that protein VII is necessary and sufficient for retention in the chromatin of members of the high-mobility-group protein B family (HMGB1, HMGB2 and HMGB3). HMGB1 is actively released in response to inflammatory stimuli and functions as a danger signal to activate immune responses[6,7]. We showed that protein VII can directly bind HMGB1 *in vitro* and further demonstrated that protein VII expression in mouse lungs is sufficient to decrease inflammation-induced HMGB1 content and neutrophil recruitment in the bronchoalveolar lavage fluid. Together, our *in vitro* and *in vivo* results show that protein VII sequesters HMGB1 and can prevent its release. This study uncovers a viral strategy in which nucleosome binding is exploited to control extracellular immune signalling.

As viruses commandeer cellular functions to promote viral production, they induce numerous cellular changes. Manipulation of host chromatin is important for viral takeover of cellular functions[1,8–11]. Although there are known examples of viral control by manipulating gene expression[2,9,12], an alternative strategy for immune evasion could exploit cellular chromatin to affect extracellular signalling. Genomes of DNA viruses are compacted and packaged into virus particles with small basic proteins encoded by the host or virus. Adenoviruses encode protein VII, a small basic protein packaged with viral genomes[3–5]. We hypothesized that protein VII contributes to host chromatin manipulation. We investigated protein VII localization during infection, and found it present in both viral replication centres stained for viral DNA-binding protein (DBP; Fig. 1a and Extended Data Fig. 1a), and in cellular chromatin stained for histone H1 and 4′,6-diamidino-2-phenylindole (DAPI; Fig. 1b). These observations suggest that protein VII functions on both viral and host genomes. To determine the impact of protein VII on cellular chromatin, we generated cell lines with inducible expression. In multiple cell types we observed that protein VII accumulation altered nuclear DNA into a punctate appearance (Fig. 1c and Extended Data Fig. 1b, c). We tested whether other basic proteins produce similar effects on chromatin. Viral core protein V and the precursor of protein VII (preVII) localized to nucleoli and did not affect chromatin appearance (Extended Data Fig. 1d). Human protamine PRM1, a basic protein involved in sperm DNA compaction[13], also localized to nucleoli and did not affect chromatin appearance (Extended Data Fig. 1d). Taken together, our data demonstrate that protein VII is sufficient to alter cellular chromatin and is distinct from other small basic proteins.

To affect cellular chromatin at the nucleosome level during infection, we reasoned that protein VII must be abundant and associated with histones. Acid extraction of histones[14,15] from infected cells revealed viral proteins VII and V isolated with cellular histones (Fig. 1d), as verified by western blot (Extended Data Fig. 2a) and mass spectrometry (MS). Protein VII abundance was comparable to cellular histone levels (Fig. 1d). We further analysed association of protein VII with cellular chromatin by salt fractionation of nuclei[16]. We found protein VII with cellular histones and DNA in high-salt fractions (Fig. 1e and Extended Data Fig. 2b–d). Ectopically expressed protein VII is also found in high-salt fractions, in contrast to other viral proteins that elute at low salt (Fig. 1e and Extended Data Fig. 2b). These data suggest that protein VII is highly abundant and tightly associated with cellular chromatin.

We hypothesized that protein VII interacts with chromatin by forming complexes with DNA, histones or nucleosomes, and examined protein VII interactions *in vitro*. Purified recombinant protein VII binds to DNA[5] (Extended Data Fig. 2e,f). We reconstituted nucleosomes *in vitro* with recombinant histone proteins on 195 base pairs (bp) of DNA[17]. Protein VII changed nucleosome mobility upon native gel electrophoresis (Fig. 1f and Extended Data Fig. 2g). We analysed native gel bands by denaturing SDS–polyacrylamide gel electrophoresis (SDS–PAGE), and confirmed that complexes contained core histones with protein VII (Fig. 1f, bottom). Unlike protamines[13], protein VII forms complexes with nucleosomes but does not appear to replace histones. Next, we examined whether protein VII association

[1]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. [2]Division of Cancer Pathobiology, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [3]Cell and Molecular Biology Graduate Group, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. [4]Department of Biochemistry and Biophysics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. [5]Epigenetics Program, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. [6]Villanova University, Villanova, Pennsylvania 19085, USA. [7]Division of Cell Pathology, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [8]Division of Pulmonary, Allergy, and Critical Care Medicine, Hospital of the University of Pennsylvania, and the Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. [9]Department of Molecular Genetics and Microbiology, School of Medicine, Stony Brook University, Stony Brook, New York 11794, USA. [10]Protein and Proteomics Core, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [11]Division of Neonatology, Children's Hospital of Philadelphia, Philadelphia, and Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. †Present address: Biotechnology Centre of Oslo and Department of Chemistry, University of Oslo, Oslo 0316, Norway.

© 2016 Macmillan Publishers Limited. All rights reserved

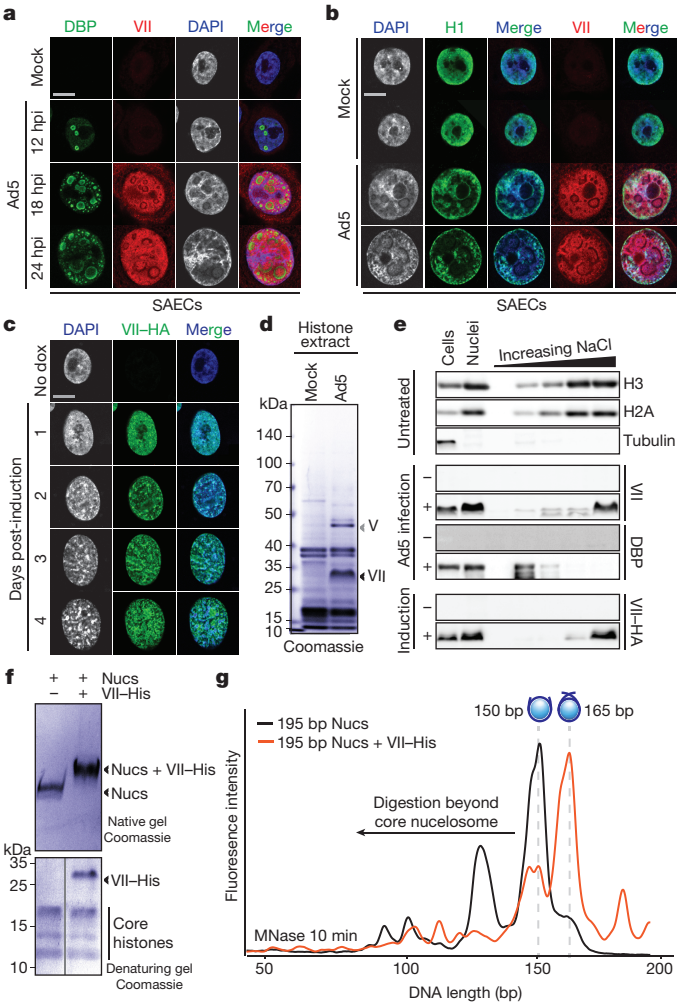


Figure 1 | Protein VII is sufficient to alter chromatin and directly binds nucleosomes. a, **b**, Adenovirus serotype 5 (Ad5)-infected small airway epithelial cells (SAECs) stained for protein VII (red) with DBP (**a**) or histone H1 (**b**) (green), and DAPI (grey, blue in merge). hpi, hours post-infection. **c**, Protein-VII–haemagglutinin (HA)-induced cells over 4 days showing HA (green) and DAPI (grey, blue in merge). dox, doxycycline. **a**–**c**, Scale bars, 10 μm. **d**, SDS–PAGE of histone extract from Ad5-infected cells showing protein V and protein VII. **e**, Western blot of chromatin fractionation from nuclei of Ad5-infected cells, induced for protein-VII–HA, or untreated. **f**, Protein VII binds to nucleosomes (Nucs). Protein bands from native gel stained with Coomassie (top) were subjected to two-dimensional analysis by SDS–PAGE (bottom). **g**, Protein VII protects nucleosome complexes from MNase digestion. Bioanalyzer curves represent nucleosomes alone (black) or protein-VII–nucleosome complexes (orange).

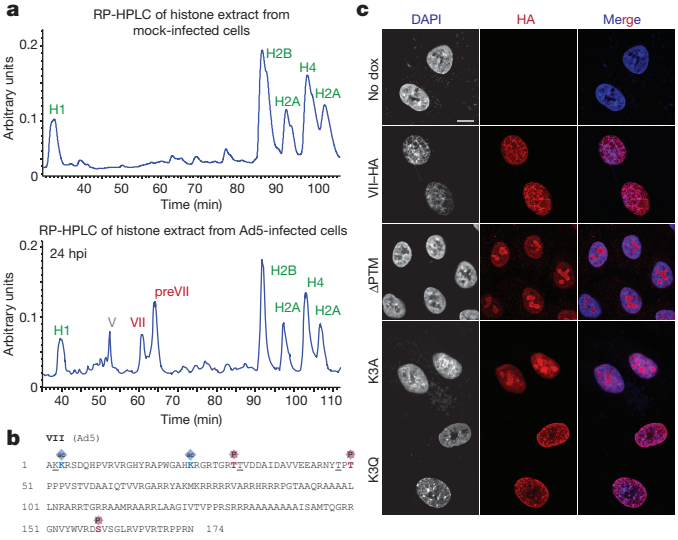


Figure 2 | Post-translational modifications on protein VII contribute to chromatin localization. a, RP-HPLC analysis of histone extracts. Viral proteins V, VII and preVII are indicated at 24 hpi. **b**, Primary sequence of protein VII with modified residues identified in infected cells. Underlined residues represent moieties that may also be modified in identified peptides (see Extended Data Fig. 5). ac, acetylated; P, phosphorylated. **c**, Immunofluoresence showing DAPI (grey, blue in merge) and protein VII (red) as wild type or with alanine substitutions at PTM sites (ΔPTM), K3A or K3Q. Scale bar, 10 μm.

with nucleosomes affects DNA wrapping using micrococcal nuclease (MNase) digestion followed by DNA fragment analysis[17]. We found that protein VII pauses nucleosomal DNA digestion at ~165 bp, the point at which DNA strands cross over the nucleosome dyad (Fig. 1g and Extended Data Fig. 3a). By contrast, nucleosome digestion alone paused with core particles at ~150 bp, suggesting that protein VII encumbers DNA access. Unlike linker histone binding that is dependent on DNA length[18], protein VII protects against MNase digestion on the nucleosome core particle of 147 bp (Extended Data Fig. 3b). Protein VII alone protects DNA from MNase digestion, as would be expected given its role in the viral core. Together, these data demonstrate that protein VII binds directly to nucleosomes and limits DNA accessibility at the DNA entry/exit site.

Post-translational modifications (PTMs) on histones are central to regulating chromatin structure[14,19]. Owing to the histone-like nature of protein VII (ref. 3), we hypothesized that it is subject to post-translational modification similar to histones. PreVII was previously proposed to be acetylated by N-terminal addition during protein synthesis[20]. We noted that protein VII contains conserved lysine residues within an AKKRS motif[21], similar to the commonly modified canonical histone motif ARSK[19]. We therefore purified protein VII from histone extracts over an adenovirus infection time course by reverse-phase high-performance liquid chromatography (RP-HPLC; Fig. 2a and Extended Data Fig. 4). Consistent with observations from histone extracts (Fig. 1d), protein VII levels were comparable to endogenous histones. We digested purified protein VII and preVII with chymotrypsin to distinguish the two proteins, and analysed peptides by tandem mass spectrometry (MS/MS). We identified several PTMs, with two acetylation sites and three phosphorylation sites the most abundant modifications (Fig. 2b and Extended Data Figs 5, 6b). Interestingly, we identified acetylation sites on ectopically expressed protein VII but not on protein VII in virus particles (Extended Data Fig. 6a). We speculate that this provides a possible mechanism for distinguishing protein VII bound to cellular chromatin from protein destined for packaged virus. To investigate the relevance of the identified PTMs, we mutated modified sites in protein VII. An alanine-replacement mutant for all five PTM sites localized to nucleoli instead of cellular chromatin (Fig. 2c). Results with individual point mutations suggest that the K3 residue is important for chromatin localization, and employing glutamine as an acetylation mimic (K3Q) mirrored the pattern of wild-type protein (Fig. 2c). Effects induced by protein VII are not due to global alteration of histone PTMs since only six PTMs on histones H3 and H4 showed minor but significant changes (Extended Data Fig. 6c–e and Supplementary Table 1). These data suggest that protein VII modification has critical roles during virus infection.

To determine whether protein VII manipulation of cellular chromatin is part of a strategy to counteract host defences, we employed MS to examine changes in the protein composition of nuclear fractions. We compared the total chromatin proteome in the presence and absence of protein VII (Fig. 3a and Supplementary Table 2). We identified 20 proteins that changed significantly across three biological replicates (Extended Data Fig. 7 and Supplementary Table 2). The categories of proteins most significantly changed upon protein VII expression were

© 2016 Macmillan Publishers Limited. All rights reserved

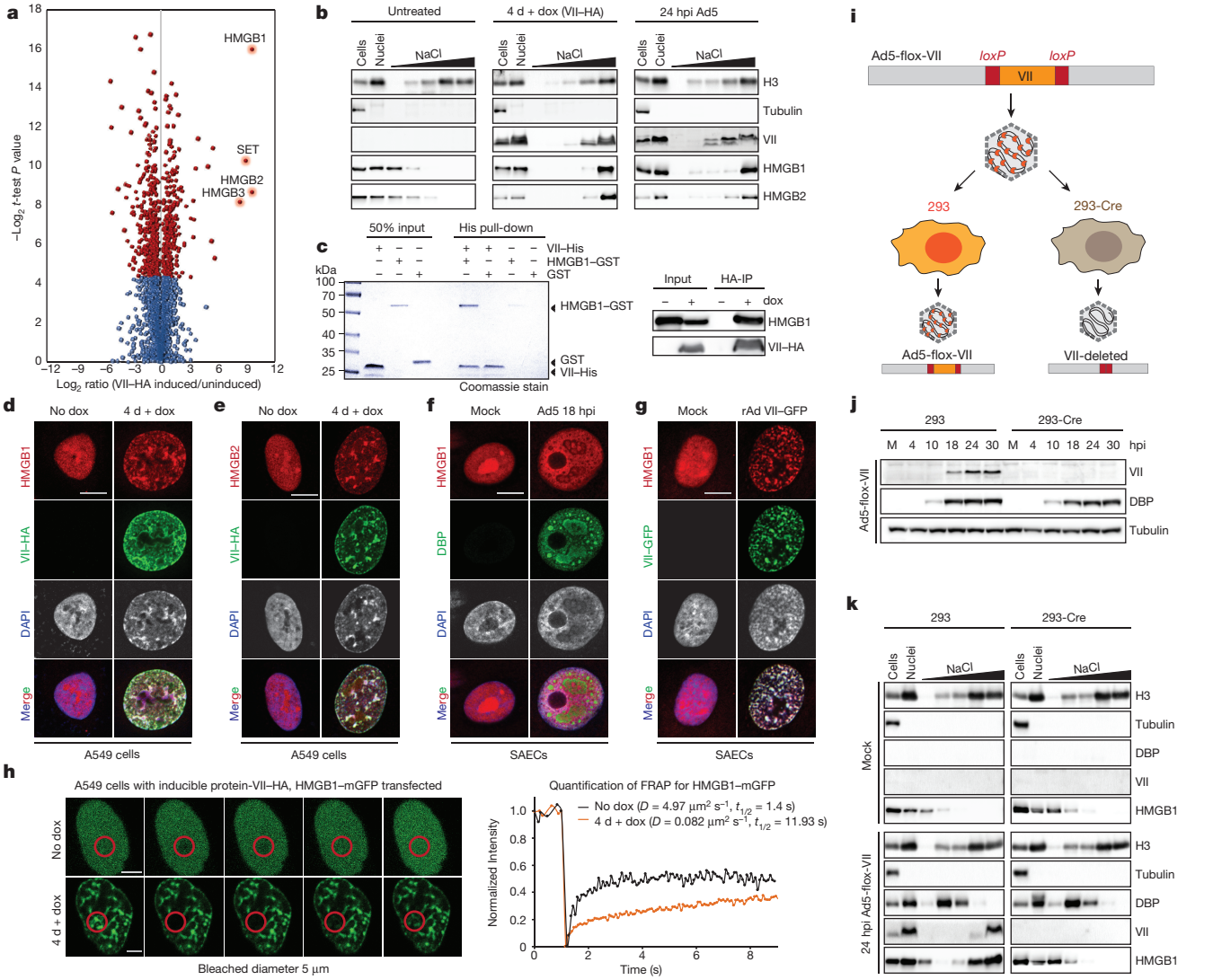


Figure 3 | Protein VII directly binds HMGB1 and is necessary for retention of the alarmin in cellular chromatin. a, Volcano plot for proteomics analysis of one representative biological replicate of the high-salt fraction. The *y* axis represents $-\log_2$ statistical *P* value and the *x* axis represents $\log_2$ protein fold-change between uninduced or protein-VII-expressing cells (homoscedastic two-tailed *t*-test, $P < 0.05$ red dots; $n = 3$ technical replicates). **b**, Nuclear fractionation shows that HMGB1 and HMGB2 normally elute from nuclei at low salt concentrations but are retained in high-salt fractions by protein-VII–HA. d, day; dox, doxycycline. **c**, Protein VII interacts with HMGB1 in pull-down of recombinant HMGB1–glutathione *S*-transferase (GST) (left, Coomassie-stained SDS–PAGE) and immunoprecipitation of HMGB1 (right, western blots). **d**, **e**, Protein VII expression alters localization of HMGB1 (**d**) and HMGB2 (**e**). Immunofluorescence shows protein-VII–HA (green) colocalized with HMGB1 (**d**) and HMGB2 (**e**) (red) in cellular chromatin (DAPI, grey, blue in merge). **f**, Same as **d** at 18 hpi with Ad5 DBP (green). **g**, Protein-VII–GFP relocalizes HMGB1 (red) to chromatin with DAPI (grey, blue in merge). rAd, recombinant adenovirus. **d**–**g**, Scale bars, 10 μm. **h**, FRAP experiment with HMGB1–monomeric GFP (mGFP). Recovery of FRAP signal in time-course images (left) with quantification and diffusion coefficients (right). Scale bar, 5 μm. *D*, diffusion coefficient; $t_{1/2}$, halftime of recovery. **i**, Schematic showing *loxP* strategy for deleting protein VII. **j**, Western blots comparing 293 and 293-Cre cells infected with Ad5-flox-VII virus. **k**, Salt fractionation in nuclei from **j**.

related to immune responses (Extended Data Fig. 7c). The top four proteins enriched in chromatin fractions by protein VII were SET (also known as TAF-1), a protein previously shown to interact with protein VII[22,23], and HMGB1, HMGB2 and HMGB3 (Fig. 3a). The HMGB proteins are alarmins with multiple functions as activators of immunity and inflammation[6,7]. HMGB1 is a nuclear protein normally only transiently associated with chromatin[24,25]. Cells also release HMGB1 as an extracellular danger signal that promotes immune responses after injury or infection[26]. We confirmed increased chromatin association of HMGB1 and HMGB2 by analysis of fractionated nuclei, upon protein VII expression and during adenovirus infection (Fig. 3b). We verified that these changes are not due to altered HMGB1 expression levels (Extended Data Fig. 8a, b). We demonstrated direct binding of recombinant protein VII to HMGB1 *in vitro* and confirmed HMGB1 co-immunoprecipitation with protein VII (Fig. 3c). We visually observed reorganization of HMGB1 and HMGB2 distribution upon protein VII expression, and at late stages of infection (Fig. 3d–f and Extended Data Fig. 8c–e). We also showed reorganization of HMGB1 distribution by vector transduction to express protein-VII–green fluorescent protein (GFP; Fig. 3g and Extended Data Fig. 8f). The effect of protein VII on HMGB1 is also conserved across human adenovirus serotypes (Extended Data Fig. 8g). We further defined the effects of protein VII on HMGB1 mobility by fluorescence recovery after photobleaching (FRAP) and found decreased HMGB1 diffusion (Fig. 3h). We next investigated whether protein VII is necessary for chromatin retention of HMGB1 during virus infection. We used a replication-competent adenovirus with *loxP* sites inserted on either side of the protein VII gene, allowing deletion of protein VII during

© 2016 Macmillan Publishers Limited. All rights reserved

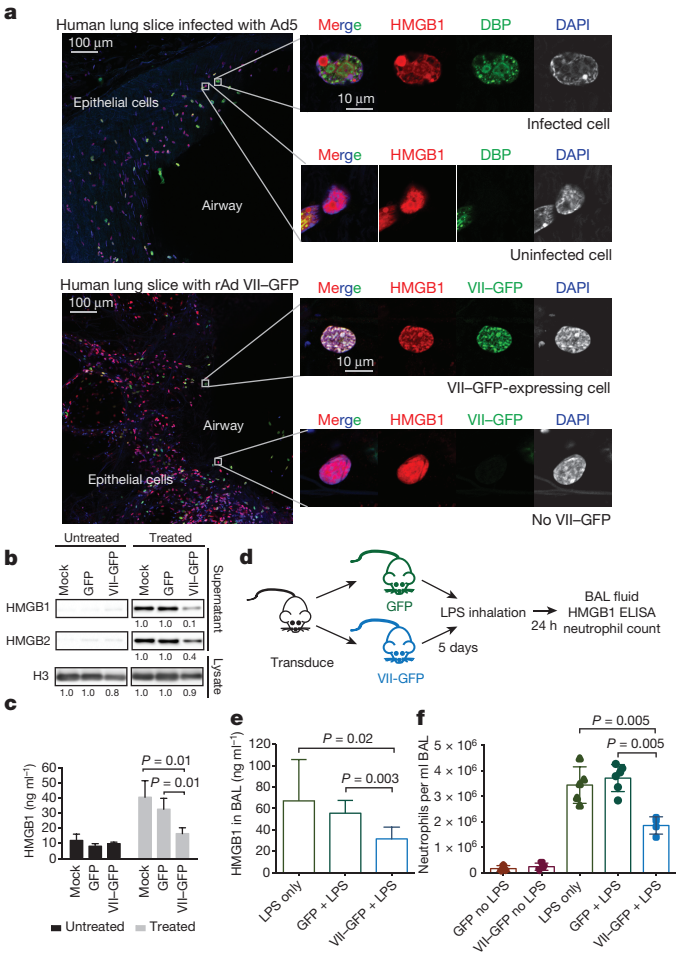


Figure 4 | Protein VII prevents HMGB1 release. a, Precision-cut lung slices infected with Ad5 or transduced to express protein-VII–GFP. Endogenous HMGB1 (red) is redistributed in cells with virus (DBP, top) and protein-VII–GFP (bottom). **b**, Protein-VII–GFP is sufficient to inhibit HMGB1 and HMGB2 release in THP-1 cells. Numbers indicate relative intensities of bands quantified with ImageJ. **c**, Enzyme-linked immunosorbent assay (ELISA)-based quantification of HMGB1 in supernatants from **b**. Mean ± standard deviation (s.d.), $n = 4$ technical replicates, homoscedastic one-tailed t-test. **d**, Schematic for investigating protein VII in a mouse lung injury model. **e**, Expression of protein-VII–GFP decreases HMGB1 in mouse BAL fluid as quantified by ELISA. Mean ± s.d., biological replicates: $n_{LPS} = 4$, $n_{GFP+LPS} = 6$, $n_{VII-GFP+LPS} = 7$, homoscedastic one-tailed ($P = 0.02$) or two-tailed ($P = 0.003$) t-test. **f**, Neutrophils in bronchoalveolar lavage (BAL) fluid are significantly fewer in mice expressing protein-VII–GFP. Mean ± s.d., biological replicates: $n_{GFP+LPS} = 6$, $n_{VII-GFP+LPS} = 4$, $n_{LPS} = 5$, $n_{GFP} = 3$, $n_{VII-GFP} = 3$, homoscedastic two-tailed t-test.

infection of cells expressing Cre recombinase (Fig. 3i, j and Extended Data Fig. 9a. b). We fractionated nuclei from infected cells and found that HMGB1 and HMGB2 were no longer retained in chromatin when protein VII was deleted (Fig. 3k and Extended Data Fig. 9c). Together, these data indicate that protein VII is necessary and sufficient to promote chromatin association and immobilization of HMGB1.

We hypothesized that protein VII retains HMGB1 in chromatin during natural infection to prevent cellular release and abrogate host immune responses. We therefore visualized endogenous HMGB1 during adenovirus infection in precision-cut lung slices[27] from human donors (Fig. 4a). Consistent with cell culture experiments, we demonstrate that protein VII is sufficient to relocalize endogenous HMGB1. We then tested whether protein VII prevents HMGB1 release in cell culture and *in vivo* models. We expressed GFP or protein-VII–GFP in macrophage-like THP-1 cells, and confirmed that protein-VII–GFP was sufficient to alter chromatin and HMGB1 localization (Extended Data Fig. 9d). Cells were treated to stimulate inflammasomes, and HMGB1 content was analysed in supernatants. Protein VII expression resulted in reduced levels of HMGB1 and HMGB2 in supernatants (Fig. 4b, c). Subsequently, we employed a murine model of lipopolysaccharide (LPS)-induced lung injury[28] to investigate the impact of protein VII on HMGB1 release and neutrophil recruitment *in vivo* (Fig. 4d). We confirmed that protein VII was expressed in transduced mouse lungs (Extended Data Fig. 10a–c) and retained mouse HMGB1 (Extended Data Fig. 9e, f). We exposed mice to inhaled LPS to induce HMGB1 release and neutrophil recruitment to alveoli. Bronchoalveolar lavage fluid obtained 24 h after LPS exposure showed that mice transduced to express protein VII had significantly less HMGB1 and fewer neutrophils than mice expressing GFP (Fig. 4d–f). Together, these data suggest that protein VII functions in cellular chromatin to retain HMGB1 as a mechanism to blunt immune responses.

In addition to known roles on packaged viral DNA[29,30], we show that protein VII interacts with cellular chromatin and binds nucleosomes. We suggest that protein VII PTMs contribute to chromatin localization, and that protein VII affects the chromatin association of host proteins. Finally, we show that protein VII in cellular chromatin leads to sequestration of HMGB family members, contributing to abrogated immune responses (Extended Data Fig. 10d). Our study reveals that chromatin retention of signalling molecules by a viral protein may represent a previously unrecognized immune evasion strategy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 March; accepted 11 May 2016.
Published online 29 June 2016.

1. Elde, N. C. & Malik, H. S. The evolutionary conundrum of pathogen mimicry. *Nature Rev. Microbiol.* **7,** 787–797 (2009).
2. Smale, S. T., Tarakhovsky, A. & Natoli, G. Chromatin contributions to the regulation of innate immunity. *Annu. Rev. Immunol.* **32,** 489–511 (2014).
3. Lischwe, M. A. & Sung, M. T. A histone-like protein from adenovirus chromatin. *Nature* **267,** 552–554 (1977).
4. Chatterjee, P. K., Vayda, M. E. & Flint, S. J. Adenoviral protein VII packages intracellular viral DNA throughout the early phase of infection. *EMBO J.* **5,** 1633–1644 (1986).
5. Vayda, M. E., Rogers, A. E. & Flint, S. J. The structure of nucleoprotein cores released from adenovirions. *Nucleic Acids Res.* **11,** 441–460 (1983).
6. Kang, R. *et al.* HMGB1 in health and disease. *Mol. Aspects Med.* **40,** 1–116 (2014).
7. Lotze, M. T. & Tracey, K. J. High-mobility group box 1 protein (HMGB1): nuclear weapon in the immune arsenal. *Nature Rev. Immunol.* **5,** 331–342 (2005).
8. Paschos, K. & Allday, M. J. Epigenetic reprogramming of host genes in viral and microbial pathogenesis. *Trends Microbiol.* **18,** 439–447 (2010).
9. Marazzi, I. *et al.* Suppression of the antiviral response by an influenza histone mimic. *Nature* **483,** 428–433 (2012).
10. Ferrari, R., Berk, A. J. & Kurdistani, S. K. Viral manipulation of the host epigenome for oncogenic transformation. *Nature Rev. Genet.* **10,** 290–294 (2009).
11. Knipe, D. M. *et al.* Snapshots: chromatin control of viral infection. *Virology* **435,** 141–156 (2013).
12. Ferrari, R. *et al.* Adenovirus small E1A employs the lysine acetylases p300/CBP and tumor suppressor Rb to repress select host genes and promote productive virus infection. *Cell Host Microbe* **16,** 663–676 (2014).
13. Wykes, S. M. & Krawetz, S. A. The structural organization of sperm chromatin. *J. Biol. Chem.* **278,** 29471–29477 (2003).
14. Lin, S. & Garcia, B. A. Examining histone posttranslational modification patterns by high-resolution mass spectrometry. *Methods Enzymol.* **512,** 3–28 (2012).
15. Shechter, D., Dormann, H. L., Allis, C. D. & Hake, S. B. Extraction, purification and analysis of histones. *Nature Protocols* **2,** 1445–1457 (2007).
16. Teves, S. S. & Henikoff, S. Salt fractionation of nucleosomes for genome-wide profiling. *Methods Mol. Biol.* **833,** 421–432 (2012).
17. Falk, S. J. *et al.* Chromosomes. CENP-C reshapes and stabilizes CENP-A nucleosomes at the centromere. *Science* **348,** 699–703 (2015).
18. White, A. E., Hieb, A. R. & Luger, K. A quantitative investigation of linker histone interactions with nucleosomes and chromatin. *Sci. Rep.* **6,** 19122 (2016).
19. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
20. Fedor, M. J. & Daniell, E. Acetylation of histone-like proteins of adenovirus type 5. *J. Virol.* **35,** 637–643 (1980).

© 2016 Macmillan Publishers Limited. All rights reserved

21. Robinson, C. M. *et al.* Molecular evolution of human adenoviruses. *Sci. Rep.* **3,** 1812 (2013).
22. Gyurcsik, B., Haruki, H., Takahashi, T., Mihara, H. & Nagata, K. Binding modes of the precursor of adenovirus major core protein VII to DNA and template activating factor I: implication for the mechanism of remodeling of the adenovirus chromatin. *Biochemistry* **45,** 303–313 (2006).
23. Haruki, H., Okuwaki, M., Miyagishi, M., Taira, K. & Nagata, K. Involvement of template-activating factor I/SET in transcription of adenovirus early genes as a positive-acting factor. *J. Virol.* **80,** 794–801 (2006).
24. Scaffidi, P., Misteli, T. & Bianchi, M. E. Release of chromatin protein HMGB1 by necrotic cells triggers inflammation. *Nature* **418,** 191–195 (2002).
25. Sapojnikova, N. *et al.* Biochemical observation of the rapid mobility of nuclear HMGB1. *Biochim. Biophys. Acta* **1729,** 57–63 (2005).
26. Lu, B. *et al.* Novel role of PKR in inflammasome activation and HMGB1 release. *Nature* **488,** 670–674 (2012).
27. Koziol-White, C. J., Damera, G. & Panettieri, R. A. Jr. Targeting airway smooth muscle in airways diseases: an old concept with new twists. *Expert Rev. Respir. Med.* **5,** 767–777 (2011).
28. Ueno, H. *et al.* Contributions of high mobility group box protein in experimental and clinical acute lung injury. *Am. J. Respir. Crit. Care Med.* **170,** 1310–1316 (2004).
29. Johnson, J. S. *et al.* Adenovirus protein VII condenses DNA, represses transcription, and associates with transcriptional activator E1A. *J. Virol.* **78,** 6459–6468 (2004).
30. Karen, K. A. & Hearing, P. Adenovirus core protein VII protects the viral genome from a DNA damage response at early times after infection. *J. Virol.* **85,** 4135–4142 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Weitzman laboratory for insightful discussions and input, especially R. Dilley and B. Simpson for generating reagents. We also thank R. Panetierri and C. Koziol-White for providing precision-cut lung slices. We are grateful to D. Curiel for sharing recombinant protein-VII–GFP vectors and L. Gerace for anti-protein-VII antibodies. We thank the Penn Vector Core for assistance in purifying recombinant vectors, the Penn CDB Microscopy Core for imaging and FRAP assistance, and the CHOP Pathology core for immunostaining of mouse lungs. We thank members of the Black, Garcia and Worthen laboratories for technical help. We thank C. Bassing, I. Brodsky, J. Henao-Mejia, R. Kohli, C. Lopez, A. Resnick, S. Shin, K. Spindler and J. Weitzman for advice and critical reading of the manuscript. D.C.A. was supported in part by T32 CA115299 and F32 GM112414. N.J.P. was supported in part by T32 NS007180. N.S. was supported in part by funding from the American Cancer Society. Research was supported by grants from the National Institutes of Health (CA097093 to M.D.W., AI102577 and CA122677 to P.H., AI118891 and GM110174 to B.A.G., and GM082989 to B.E.B.), the Institute for Immunology of the University of Pennsylvania, and funds from the Children's Hospital of Philadelphia (M.D.W.).

Author Contributions D.C.A. and M.D.W. conceived the project and designed experiments; D.C.A., C.H., N.S., J.P., N.J.P. and E.D.R. performed the experiments; D.C.A., C.H. and J.P. generated constructs and cell lines; K.K., R.C.M., S.H.S. and B.A.G. performed MS analysis; P.O. and P.H. generated Ad5-flox-VII virus and provided 293-Cre cell line; D.C.A. and D.B. performed the FRAP experiments; A.J.P. and G.S.W. conducted all mouse experiments; B.E.B. and B.A.G. designed experiments and interpreted the data; D.C.A. and M.D.W. interpreted the data and wrote the manuscript and all authors were involved in editing the manuscript.

Author Information All proteomic raw files have been deposited in the Chorus database under project number 1047 (https://chorusproject.org/). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.D.W. (weitzmanm@email.chop.edu).

Reviewer Information *Nature* thanks M. Bianchi and the other anonymous reviewer(s) for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Cells. Primary SAECs, U2OS, HeLa, 293, THP-1 and A549 cells were obtained from the American Type Culture Collection (ATCC) and grown according to the provider's instructions. Cell lines were not authenticated or tested for mycoplasma. Acceptor cells for generation of inducible cell lines were provided by E. Makeyev and used as previously reported[31]. Protein VII, preVII and V were cloned from genomic DNA isolated from HeLa cells infected with adenovirus type 5 and inserted into the inducible plasmid cassette with a C-terminal HA tag using restriction enzymes BsrGI and AgeI (primer sequences available upon request). Positive clones were selected in DH5α cells, sequenced, and transfected into A549, U2OS or HeLa acceptor cells along with plasmid expressing the Cre recombinase. Recombined clones were selected by puromycin resistance ($1\,\mu g\,ml^{-1}$) and induced with doxycycline ($0.2\,\mu g\,ml^{-1}$) to express the desired protein. Protein expression was verified by immunofluorescence and western blot. All figures shown are after 4 days of induction unless otherwise stated. Protein VII and preVII were also verified by HPLC purification and MS analysis. Point mutations were generated by gene synthesis from Genewiz. 293-Cre cells were provided by P. Hearing.

Viruses and infections. Wild-type Ad5, Ad9, Ad12 and recombinant adenovirus vectors expressing only GFP were propagated in 293 cells as previously described[32]. Recombinant adenovirus vector with protein-VII–GFP replaced in the E1 region was a gift from D. Curiel[33]. Infections were carried out as described previously[34] using a multiplicity of infection of 10 for primary cells and cell lines for Ad5 infections. Ad9 and Ad12 infections were carried out with a multiplicity of infection of 50 and 20, respectively. Ad5-flox-VII was generated by P. Hearing and also prepared using standard methods in 293 cells. *loxP* sites were added flanking protein VII in the Ad5 genome resulting in protein VII deletion during infection of 293 cells expressing Cre recombinase.

Antibodies. Primary antibodies were purchased from Covance (HA MMS-101R), Abcam (H1 ab4269, H3 ab1791, HMGB1 ab18256, HMGB2 ab67282), Millipore (H2A 07-146, prosurfactin-C AB3786), and Santa Cruz (Ku86 sc5280, tubulin sc69969). The antibodies to DBP, adenoviral late proteins, terminal protein and protein VII were gifts from A. Levine[35], J. Wilson[32], R. Hay and L. Gerace, respectively. Secondary antibodies for immunoblotting were obtained from Jackson ImmunoResearch and secondary antibodies for immunofluorescence were obtained from Life Technologies.

Immunofluorescence. Cells were grown on glass coverslips in 24-well plates and either infected or induced with doxycycline ($0.2\,\mu g\,ml^{-1}$). Cells were harvested for immunofluorescence at the indicated time points, washed in PBS, fixed in 4% paraformaldehyde for 15 min and post-fixed with 100% ice-cold methanol for 5 min. Coverslips were then blocked and stained as previously described[36] and mounted using ProLong Gold Antifade Reagent (Life Technologies). Immunofluorescence was visualized using a Zeiss LSM 710 Confocal microscope (Cell and Developmental Microscopy Core at UPenn) and ZEN 2011 software. Images were processed using ImageJ and assembled with Adobe CS6.

Immunoblotting. Western blot analysis was carried out using standard methods. Briefly, equal amounts of total protein lysates were separated by SDS–PAGE and transferred to a nitrocellulose membrane (Millipore) for at least 30 min at 30 V. Membranes were stained with ponceau to confirm protein loading and blocked in 5% milk in TBST containing 0.1% azide. Membranes were incubated with primary antibodies overnight, washed for 30 min in TBST and incubated with secondary antibodies conjugated to horseradish peroxidase (Jackson Laboratories) for 1 h. Membranes were washed again and proteins were visualized with Pierce ECL Western Blotting Substrate (Thermo Scientific) and detected using a Syngene G-Box.

Mice. All mice were housed in specific-pathogen-free (SPF) conditions in an animal facility at the Children's Hospital of Philadelphia. All studies in mice were carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and approved by the Institutional Animal Care and Use Committee, Children's Hospital of Philadelphia Animal Welfare Assurance Number A3442-01. C57BL/6J male mice aged 8–10 weeks were used for experiments. Mice were sedated with ketamine and xylazine. Once sedated, mice underwent orotracheal intubation, as previously described[37], with a 20G angiocatheter from BD. Mice subsequently received 5×10^{10} genome copies (GC) of recombinant adenovirus expressing protein-VII–GFP or GFP purified by the Penn Vector Core. Four days after infection, mice were exposed to aerosolized LPS, $3\,mg\,ml^{-1}$ for 30 min as previously described[38]. One day after LPS exposure, bronchoalveolar lavage (BAL) and lung tissue were harvested as previously detailed[39] and examined for HMGB1 content (ELISA, Chondrex 6010) and neutrophil count (haematoxylin and eosin stain kit EMD 65044/93). Immunostaining was carried out by the CHOP Pathology Core using standard methods. A minimum of four biological replicates were used for each condition studied. Mice were assigned a random number and colour at the start of the experiment and were randomized. Technicians carrying out the experiments were blinded to the identity of the samples. Tissue samples were assigned a random study number such that the technician performing the analysis was blinded. Unblinding for the purpose of data analysis occurred only after all data had been collected.

Salt fractionation of nuclei. Salt fractionation of nuclei was adapted from established protocols[16,40]. Briefly, $2–4 \times 10^7$ cells were collected and resuspended in 2 ml of ice-cold buffer I (0.32 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 15 mM Tris, pH 7.5, 0.5 mM dithiothreitol (DTT), 0.1 mM PMSF and protease inhibitor cocktail from Roche). To dissolve the plasma membrane, 2 ml ice-cold buffer I supplemented with 0.1% IGEPAL were added and samples were incubated on ice for 10 min. The 4 ml of nuclei was layered on 8 ml of ice-cold buffer II (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 15 mM Tris, pH 7.5, 0.5 mM DTT, 0.1 mM PMSF and protease inhibitor cocktail from Roche) and centrifuged for 20 min at 10,000*g* and 4 °C. The pelleted nuclei were resuspended in 400 μl buffer III (10 mM Tris pH 7.4, 2 mM $MgCl_2$, 0.1 mM PMSF) supplemented with 5 mM $CaCl_2$ and the DNA was digested to mononucleosomes by addition of 1 unit of MNase (Sigma-Aldrich, N3755). The reaction was incubated at 37 °C for 30 min and then stopped by addition of 25 μl of 0.1 M EGTA. The samples were centrifuged for 10 min, 350*g*, at 4 °C, and supernatants were set aside for western blot analysis. The pellet was resuspended in 400 μl of buffer IV (70 mM NaCl, 10 mM Tris pH 7.4, 2 mM $MgCl_2$, 2 mM EGTA, 0.1% Triton X-100, 0.1 mM PMSF) with 80 mM salt and rotated for 30 min at 4 °C. The sample was centrifuged for 10 min at 350*g*, 4 °C, and the supernatant collected for western blot analysis. This step was repeated for salt concentrations in buffer IV of 150 mM, 300 mM and 600 mM. The final pellet was resuspended in 400 μl H_2O and all samples were analysed together by western blot. An aliquot of each supernatant was set aside for DNA purification using a PCR purification kit (Qiagen) and analysed by agarose gel electrophoresis. Alternatively, 4×10^7 cells were resuspended in 400 μl hypotonic buffer (10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 1:1,000 PMSF, 0.5 mM DTT) and incubated on ice for 30 min. The cells were transferred to a 1 ml dounce tissue grinder and the cell membranes were gently disrupted with 40 strokes of a tight-fitting pestle. The samples were centrifuged for 5 min at 1,500*g* and 4 °C. The pelleted nuclei were resuspended in 400 μl buffer III and the fractionation was continued as described earlier.

Preparation of salt fractions for MS analysis. All chemicals used for preparation of MS samples were of at least sequencing grade and purchased from Sigma-Aldrich, unless otherwise stated. Only the 600 mM salt fraction was used for LC-MS/MS analysis. The 0.1% Triton X-100 detergent was removed from samples bofore MS analysis by precipitation using chloroform ($CHCl_3$)-methanol (MeOH) precipitation[41]. The protein pellet from $CHCl_3$-MeOH precipitation was resuspended in 6 M urea and 2 M thiourea in 50 mM ammonium bicarbonate. Samples were reduced with 10 mM DTT for 1 h at room temperature and then carbamidomethylated with 20 mM iodoacetamide for 30 min at room temperature in the dark. Afterwards, alkylated proteins were digested first with endopeptidase Lys-C (Wako, MS grade) for 3 h, after which the solution was diluted 10 times with 20 mM ammonium bicarbonate. Subsequently, samples were digested with trypsin (Promega) at an enzyme-to-substrate ratio of approximately 1:50 for 12 h at room temperature. The samples were acidified with 5% formic acid (FA) to a $pH \leq 3$ and desalted using Poros Oligo R3 RP columns (PerSeptive Biosystems) packed in a P200 stage tip with C_{18} 3 M plug (3M Bioanalytical Technologies). Purified peptide samples were dried by lyophilization and stored at −20 °C until further analysis. This procedure was carried out for three biological replicas.

Nano-LC-MS/MS and analysis of salt fractions. Samples were loaded onto a 16 cm C_{18}–AQ column (inner diameter 75 μm, 3 μm beads, Dr, Maisch GmbH, Germany) using an Easy nano-flow HPLC system (Thermo Fisher Scientific). The nano-LC was coupled to an Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific) via a nanoelectrospray ion source (Thermo Fisher Scientific). Peptides were loaded in buffer A (0.1% formic acid) and eluted with a 120 min linear gradient from 2–30% buffer B (95% acetonitrile, 0.1% formic acid). After the gradient, the column was washed with 90% buffer B. Mass spectra were acquired using a data-dependent acquisition method with the TopSpeed set with 3-s cycle. Spectra were acquired in the Orbitrap analyser with mass range of 350–1,200 m/z and 120,000 resolution (200 m/z), with a maximum injection time of 50 ms and an AGC target of 5×10^5. Signals with 2–5 charges were selected for HCD fragmentation using a normalized collision energy of 27, a maximum injection time of 120 ms and an AGC target of 10,000. Fragments were analysed in the ion trap. Raw MS files were analysed by MaxQuant (v.1.5.2.8)[42] (http://www.maxquant.org). MS/MS spectra were searched against the UniProt-human database (version June 2014, 59,345 entries). All used search parameters were default, with the exception of including the match between runs (1 min window) and the intensity-based absolute quantification (iBAQ) label-free quantification[43]. The search included variable modifications of methionine oxidation and

© 2016 Macmillan Publishers Limited. All rights reserved

N-terminal acetylation, and fixed modification of carbamidomethyl cysteine. Each iBAQ value was $\log_2$ transformed and subsequently normalized by the average protein abundance within each run. Biological process association analysis and process network enrichment were performed using the GeneGo MetaCore pathways analysis package with FDR < 5%; each Gene Ontology term was ranked using *P*-value enrichment.

Purification of recombinant protein-VII–His. Protein VII was cloned from genomic DNA isolated from adenovirus-infected HeLa cells into a pET21a backbone to generate a C-terminal hexahistidine tag. Positive clones were selected in DH5α cells, sequenced, and transformed into BL21 (DE3) cells (NEB C2527I). The purification of insoluble protein-VII–His was adapted from existing protocols to purify histone proteins from *Escherichia coli*[44,45]. Briefly, BL21 cells were inoculated from overnight cultures and grown to an optical density of 0.5–0.6 $OD_{260\,nm}$, induced with 0.1 mM isopropyl-β-D-thiogalactoside (IPTG; Sigma) and harvested after 4 h at 37 °C. Cell pellets were resuspended in a mild buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM PMSF, 5% glycerol, 2.5 μg ml^{-1} aprotinin, leupeptin and pepstatin) and disrupted by sonication using a Branson 250 sonifier. The lysate was then centrifuged at 27,000*g* for 20 min at 4 °C. The supernatants were discarded and pellets were resuspended in a denaturing buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 8 M urea). The suspension was centrifuged again to eliminate insoluble cell debris and the His-tagged protein was isolated using a cobalt resin (ThermoScientific 89964) according to the manufacturer's instructions for denaturing conditions. The purified protein was then dialysed against water and lyophilized. Purified protein was verified by western blot and MS.

***In vitro* binding assays.** HMGB1–GST (Abnova) or GST (Sigma) were combined with recombinant protein-VII–His at equimolar ratios and incubated at 4 °C for 1 h. Complexes were then mixed with a cobalt resin (ThermoScientific 89964) to bind protein-VII–His and any associated protein and washed three times in the binding buffer (50 mM Tris pH 8, 300 mM NaCl, 0.1% IGEPAL). The beads were then boiled in sample buffer, separated on a 4–12% NuPage gel and visualized by Coomassie staining.

Nucleosome *in vitro* binding and MNase digestion assays. Gel shift and MNase digestion assays were carried out as previously described[17,46,47]. Briefly, nucleosomes were reconstituted by incubating purified recombinant histones with '601' DNA of either 195 or 147 bp over a series of dialysis. Recombinant protein-VII–His was then combined with nucleosomes at various molar ratios, incubated at room temperature for 15 min, and analysed by native gel electrophoresis. Complexes were also digested with MNase (Affymetrix) by addition of 1 unit per μg of DNA for 147 bp nucleosome experiments and 0.1 unit per μg of DNA for 195 bp nucleosome experiments, incubated at 22 °C for varying amounts of time followed by the addition of EGTA and guanidine thiocyanate to stop the reaction. The DNA fragments were then purified using a MinElute PCR purification kit (Qiagen) and analysed on an Agilent 2100 Bioanalyzer as previously described[17].

Release assay of HMGB1 in THP-1 cells. THP-1 cells were seeded at a density of 2×10^5 cells per well in a 24-well plate, and stimulated into macrophage-like cells by addition of 10 ng ml^{-1} PMA for 48 h. Cells were washed in PBS and transduced with recombinant adenovirus vectors expressing only GFP or protein-VII–GFP such that >90% of cells were GFP positive. At 48 h after transduction, cells were washed and 200 μl of serum-free RPMI was added. To stimulate the inflammasome, LPS (Sigma-Aldrich L2880) with a final concentration of 0.5 μg ml^{-1} was added to wells and incubated for 2 h, then nigericin (Sigma-Aldrich N7143) was added with a final concentration of 10 μM for 1 h. Supernatants were collected and proteins precipitated overnight at 4 °C with a final concentration of 20% trichloroacetic acid (Sigma), washed with acetone, dried, and resuspended in 1 × LDS sample buffer with reducing agent (Invitrogen). For ELISA analysis, supernatants were harvested directly and HMGB1 content was detected by the manufacturer's instructions (Chondrex 6010). Cells were also harvested by the addition of 1 × LDS sample buffer with reducing agent (Invitrogen) and boiled. Supernatants and lysates were analysed together by western blot.

Acid extraction and RP-HPLC. Histones were prepared for MS analysis as detailed previously[48]. Nuclei were isolated and histones from infected cells were extracted by acid as previously described[14]. The preVII and protein VII variants were fractionated using an offline RP-HPLC. Briefly, ~100 μg proteins were resuspended in buffer A (0.1% trifluoroacetic acid (TFA) in HPLC-grade water) and loaded onto a C_{18} 5 μm column (4.6 mm internal diameter × 250 mm, Vydac) using a Beckman Coulter (System GoldA) HPLC (buffer A: 0.1% TFA; buffer B: 95% acetonitrile, 0.08% TFA). The proteins were separated using a gradient from 30 to 45% buffer B in 100 min at a flow rate of 0.2 ml min^{-1}. The fractions containing the proteins of interest were collected using an automatic fraction collector and individual peaks were combined based on their ultraviolet signal. The fractions were subsequently dried by vacuum centrifugation and prepared for MS (see later). Protein VII was purified from three biological replicates and analysed as follows for MS.

MS analysis of protein VII PTMs. *Sample preparation/protein VII.* RP-HPLC-purified samples of protein VII variants were reduced in 10 mM DTT in 50 mM ammonium bicarbonate for 1 h at 56 °C. After cooling to room temperature, samples were alkylated in 20 mM iodoacetamide in 50 mM ammonium bicarbonate for 30 min in the dark. Samples were digested with chymotrypsin or Arg-C, at an enzyme-to-substrate ratio of approximately 1:20 for 8 h at 37 °C. The samples were acidified to a final concentration of 5% formic acid to a pH ≤ 3 and desalted using P200 stage tip columns packed with C_{18} 3 M plug (3M Bioanalytical Technologies). Purified peptide samples were dried by lyophilization and stored at −20 °C until further analysis.

Nano-LC-MS/MS analysis of histone PTMs. The nano-LC-MS/MS analysis was performed as previously described[48].

Nano-LC-MS/MS analysis of protein VII peptides. The nano-LC-MS/MS analysis was performed in triplicate for each sample. Samples were loaded onto a 16 cm C_{18}–AQ column (inner diameter 75 μm, 3 μm beads, Dr, Maisch GmbH) using an Easy nano-flow HPLC system (Thermo Fisher Scientific). The nano-LC was coupled to an Orbitrap Velos Pro Mass Spectrometer (Thermo Fisher Scientific) via a nanoelectrospray ion source (Thermo Fisher Scientific). Peptides were loaded in buffer A (0.1% formic acid) and eluted with a 45 min linear gradient from 2 to 30% buffer B (95% acetonitrile, 0.1% formic acid). After the gradient, the column was washed with 90% buffer B. Mass spectra were acquired using a data-dependent acquisition method with the top 15 most intense ions. Spectra were acquired in the Orbitrap analyser with mass range of 350–1,600 *m/z* and 60,000 resolution (400 *m/z*), with a maximum injection time of 10 ms and an AGC target of 10×10^6. Signals above 1,000 count charges were selected for HCD fragmentation using normalized collision energy of 36, a maximum injection time of 100 ms and an AGC target of 50,000. Fragments were analysed in the orbitrap.

Data processing of protein VII spectra. Raw mass spectrometer files were analysed using Proteome Discoverer (v.1.4, Thermo Scientific). MS/MS spectra were converted to .mgf files and searched against the UniProt adenovirus C serotype 5 database using Mascot (v.2.5, Matrix Science). Database searching was performed with the following parameters: precursor mass tolerance 10 p.p.m.; MS/MS mass tolerance 0.05 Da; enzyme chymotrypsin (Promega) or Arg-C (Roche), with two missed cleavages allowed; fixed modification was cysteine carbamidomethylation; variable modifications were methionine oxidation, serine/threonine/tyrosine phosphorylation, lysine acetylation and methylation, asparagine and glutamine deamidation. Specifically, phosphorylation, acetylation, and methylation were searched separately, not as co-existing modifications. Peptides were filtered for <1% FDR, Mascot ion score >20 and peptide rank 1.

Co-immunoprecipitation of protein-VII–HA. A549 cells were induced to express protein VII with doxycycline for 4 days as described earlier. Approximately 4×10^7 cells were harvested and pelleted for each immunoprecipitation reaction. Cell pellets were resuspended in 500 μl of IC wash buffer with protease inhibitors (20 mM HEPES pH 7.9, 110 mM KOAc, 2 mM $MgCl_2$, 150 mM NaCl, 0.1% Tween-20, 0.1% Triton X-100) and incubated on ice for 10 min with intermittent vortexing to disrupt cells. Samples were then incubated on ice for 1 h with 5 μl of benzonase (Millipore) added to each sample to digest DNA to ~150 bp, which was confirmed by DNA isolation and agarose gel analysis. Samples were then sonicated in a Diagenode Bioruptre for 30 s on and 30 s off for five rounds at 4 °C and centrifuged at 14,000*g* for 15 min at 4 °C. Supernatants were then incubated rotating for 1 h at 4 °C with 30 μl of HA-conjugated magnetic beads (Thermo Scientific) and washed three times for 5 min in IC buffer. Isolated proteins were eluted with 100 μl of 2 mg ml^{-1} HA peptide (Thermo Scientific) for 20 min rotating at 37 °C and separated on an SDS–PAGE gel. For protein separation by SDS–PAGE the NuPAGE 1DE System was used (NuPAGE Novex 4–12% Bis-Tris 1.0 mm gels, Invitrogen). Uninduced cells were used as a negative control. The immunopreciptation was carried out in biological triplicate and pull-down of protein-VII–HA and HMGB1 was confirmed by western blotting standard techniques as described earlier.

Quantitative PCR. Genomic DNA was isolated using the PureLink Genomic DNA kit (Thermo Scientific). Quantitative PCR was performed using primers specific for viral DBP (5′-GCCATTGCGCCCAAGAAGAA and 5′-CTGTCCACGATTACCTCTGGTGAT), protein VII (5′-GCGGGT ATTGTCACTGTGC and 5′-CACCCAATACACGTTGCCC), and cellular tubulin (5′-CCAGATGCCAAGTGACAAGAC and 5′-GAGTGAGTGACAA GAGAAGCC). Values for DBP and protein VII were normalized internally to tubulin and to the 4 h time point to control for any variation in virus input. RNA was isolated using the RNeasy Mini Kit (Qiagen) and reverse transcribed using the High Capacity RNA to cDNA Kit (Applied Biosystems). Quantitative PCR was performed using primers specific for HMGB1

© 2016 Macmillan Publishers Limited. All rights reserved

(5′-TAACTAAACATGGGCAAAGGAG and 5′-TAGCAGACATGGTCTTCCAC) and β-actin (5′-GCACCACACCTTCTACAATGAG and 5′-GGTCTCAA ACATGATCTGGGTC). Quantitative PCR was performed using the standard protocol for Sybr Green (Thermo Scientific) and analysed using the ViiA 7 Real-Time PCR System (Thermo Scientific).

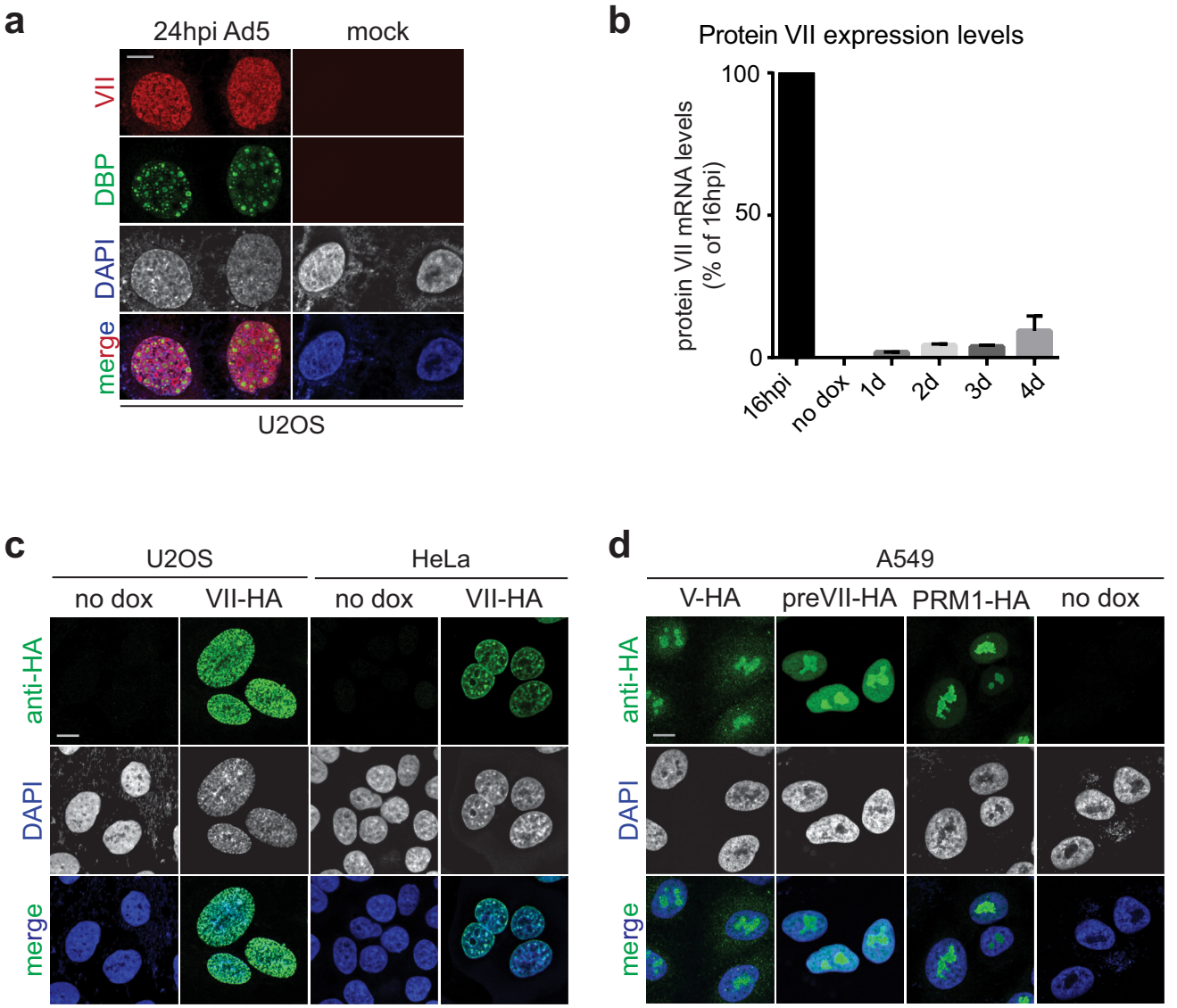
Precision-cut lung slice immunofluorescence. Precision-cut lung slices were obtained and prepared as previously described[27,49]. De-identified human lung tissue from donors was obtained from the National Disease Research Interchange. Analysis of human samples was approved by the University of Pennsylvania Internal Review Board. Samples were infected with 10^8 plaque-forming units (p.f.u.) of Ad5 per slice or 10^9 GC of rAd protein-VII–GFP for 24 h. Samples were fixed in 4% PFA at room temperature for 15 min and washed three times in PBS. Samples were permeabilized with 0.5% Triton X-100 and washed twice more in PBS. Samples were then incubated with 3% BSA and 0.03% Triton X-100 in PBS for 1 h to block. Primary antibodies (DBP or HMGB1) were incubated in the same buffer for 1 h and then samples were washed three times in PBS with 3% BSA, incubated with secondary antibodies and DAPI for 1 h, and washed three more times. Whole slices were mounted on slides with mounting solution and imaged by confocal microscopy.

FRAP. Full-length HMGB1 was cloned from pcDNA3.1 Flag-hHMGB1 (Addgene 31609) into pEGFP-N1 containing a L221K mutation to prevent dimerization of GFP molecules[50]. A549 cells were induced to express protein VII for 4 days with doxycycline in glass-bottom dishes. Cells were then transfected with the construct that constitutively expresses HMGB1 with a monomeric GFP C-terminal tag. FRAP was carried out using standard methods on a Zeiss LSM 710 confocal microscope. Diffusion coefficients were calculated using the 'simFRAP' algorithm (http://imagej.nih.gov/ij/plugins/sim-frap/index.html), a simulation based approach to FRAP analysis[51].

Statistical analyses. Statistical details are reported in each figure legend. Statistical analyses were performed on at least three different biological replicates, unless otherwise stated in the figure legend. The sample size was chosen to provide enough statistical power to apply parametric tests (one- or two-tailed homoscedastic *t*-test). The *t*-test was considered a valuable statistical test since binary comparisons were performed and the number of replicates was limited. Furthermore, we applied the homoscedastic *t*-test assuming that the variance between the two data sets would remain homogeneous due to the use of the same cell lines in culture with and without protein VII expression. No samples were excluded as outliers (this applies to all proteomics analyses described in this manuscript). Proteins with a *P* value smaller than 0.05 were considered to be significantly altered between the two tested conditions for two-tailed and one-tailed *t*-test. Data distribution was assumed to be normal but this was not formally tested. The nano-LC-MS/MS analysis was performed in triplicate for each sample to determine technical variation.

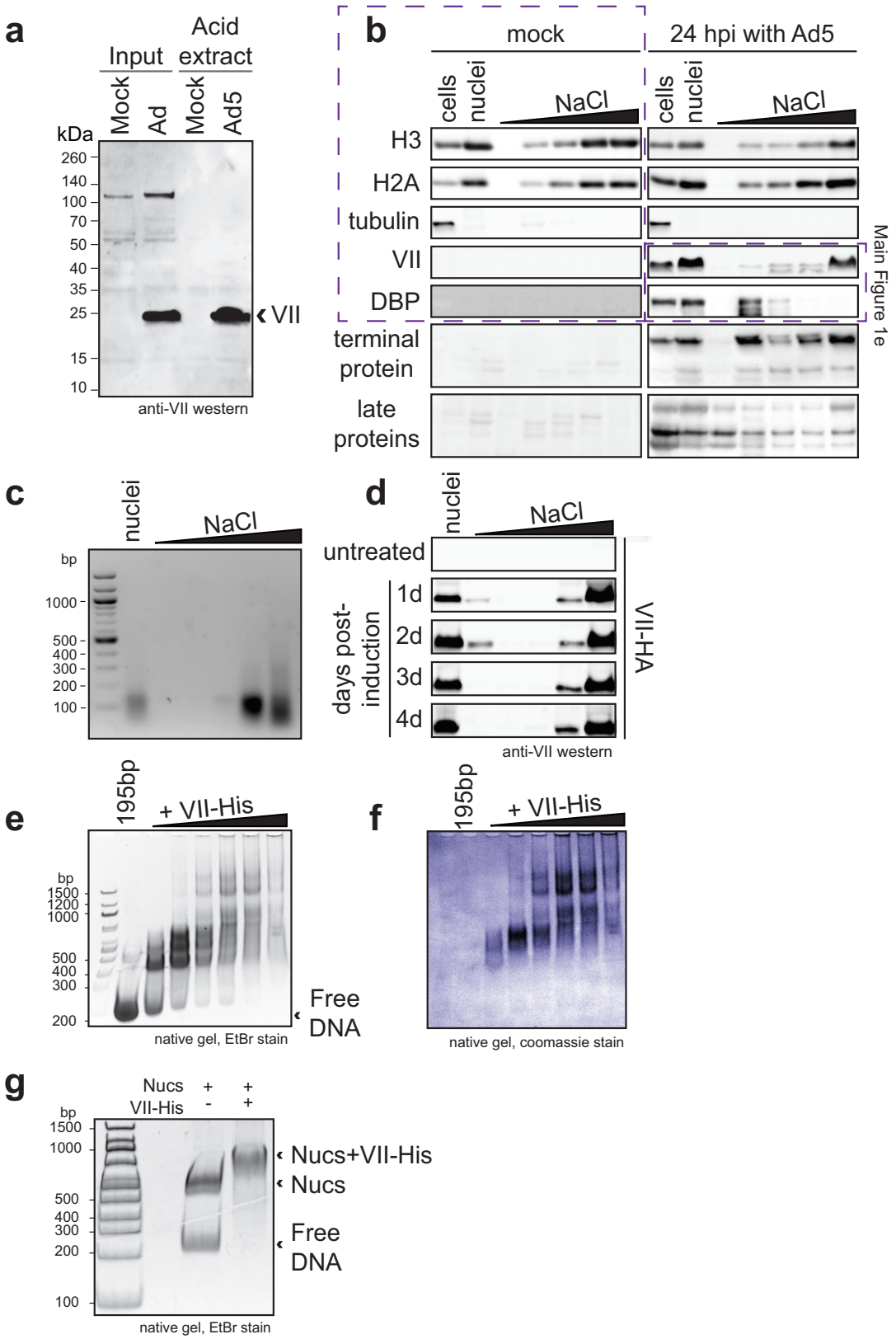
31. Khandelia, P., Yap, K. & Makeyev, E. V. Streamlined platform for short hairpin RNA interference and transgenesis in cultured mammalian cells. *Proc. Natl Acad. Sci. USA* **108,** 12799–12804 (2011).
32. Kozarsky, K. F., Jooss, K., Donahee, M., Strauss, J. F., III & Wilson, J. M. Effective treatment of familial hypercholesterolaemia in the mouse model using adenovirus-mediated transfer of the VLDL receptor gene. *Nature Genet.* **13,** 54–62 (1996).
33. Orazio, N. I., Naeger, C. M., Karlseder, J. & Weitzman, M. D. The adenovirus E1b55K/E4orf6 complex induces degradation of the Bloom helicase during infection. *J. Virol.* **85,** 1887–1892 (2011).
34. Le, L. P. *et al.* Core labeling of adenovirus with EGFP. *Virology* **351,** 291–302 (2006).
35. Reich, N. C., Sarnow, P., Duprey, E. & Levine, A. J. Monoclonal antibodies which recognize native and denatured forms of the adenovirus DNA-binding protein. *Virology* **128,** 480–484 (1983).
36. Lilley, C. E., Chaurushiya, M. S., Boutell, C., Everett, R. D. & Weitzman, M. D. The intrinsic antiviral defense to incoming HSV-1 genomes includes specific DNA repair proteins and is counteracted by the viral protein ICP0. *PLoS Pathog.* **7,** e1002084 (2011).
37. Das, S., MacDonald, K., Chang, H.-Y. S. & Mitzner, W. A simple method of mouse lung intubation. *J. Vis. Exp.* **73,** e50318 (2013).
38. Jeyaseelan, S., Chu, H. W., Young, S. K. & Worthen, G. S. Transcriptional profiling of lipopolysaccharide-induced acute lung injury. *Infect. Immun.* **72,** 7247–7256 (2004).
39. Nick, J. A. *et al.* Role of p38 mitogen-activated protein kinase in a murine model of pulmonary inflammation. *J. Immunol.* **164,** 2151–2159 (2000).
40. Zaret, K. Micrococcal nuclease analysis of chromatin structure. *Curr. Protoc. Mol. Biol.* Chapter 21, Unit 21.1 (2005).
41. Wessel, D. & Flügge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* **138,** 141–143 (1984).
42. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **26,** 1367–1372 (2008).
43. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473,** 337–342 (2011).
44. Tanaka, Y. *et al.* Expression and purification of recombinant human histones. *Methods* **33,** 3–11 (2004).
45. Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272,** 301–311 (1997).
46. Hasson, D. *et al.* The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nature Struct. Mol. Biol.* **20,** 687–695 (2013).
47. Sekulic, N., Bassett, E. A., Rogers, D. J. & Black, B. E. The structure of $(CENP\text{-}A\text{–}H4)_2$ reveals physical features that mark centromeres. *Nature* **467,** 347–351 (2010).
48. Kulej, K., Avgousti, D. C., Weitzman, M. D. & Garcia, B. A. Characterization of histone post-translational modifications during virus infection using mass spectrometry-based proteomics. *Methods* **90,** 8–20 (2015).
49. Cooper, P. R. & Panettieri, R. A. Jr. Steroids completely reverse albuterol-induced β_2-adrenergic receptor tolerance in human small airways. *J. Allergy Clin. Immunol.* **122,** 734–740 (2008).
50. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296,** 913–916 (2002).
51. Blumenthal, D., Goldstien, L., Edidin, M. & Gheber, L. A. Universal approach to FRAP analysis of arbitrary bleaching patterns. *Sci. Rep.* **5,** 11655 (2015).

© 2016 Macmillan Publishers Limited. All rights reserved



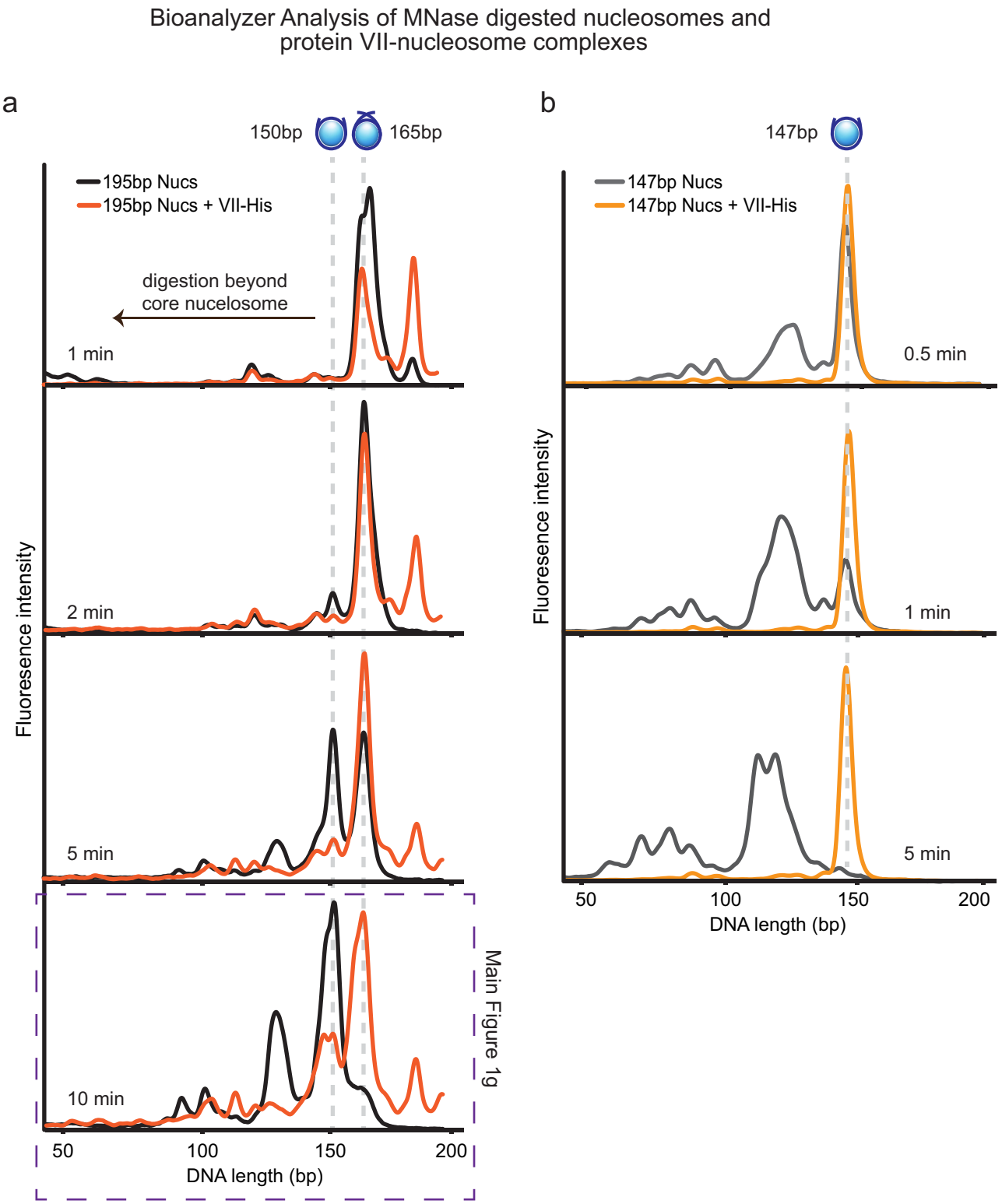
Extended Data Figure 1 | Adenovirus protein VII distorts chromatin. **a**, Protein VII localizes to cellular chromatin and viral replication centres in U2OS cells similarly to SAECs in Fig. 1a. **b**, Protein VII messenger RNA levels measured by quantitative PCR showing that after 4 days of induction in the A549 cell line, the level of protein VII transcripts is approximately 10% of that measured during infection at 16 hpi. Despite the low relative level, this amount of protein VII is sufficient to cause dramatic changes in the nucleus (graph shows mean ± s.d., $n = 3$ biological replicates). **c**, Inducible cell lines of U2OS and HeLa expressing protein-VII–HA show chromatin localization and distortion, similar to A549 cells in Fig. 1c. **d**, Inducible A549 cell lines expressing viral protein V, the precursor for protein VII (preVII) or cellular protamine PRM1 with C-terminal HA tags. Although all three proteins possess a large number of charged residues, none are sufficient to distort cellular chromatin or increase nuclear size as observed with mature protein VII. Scale bars, 10 μm.

© 2016 Macmillan Publishers Limited. All rights reserved



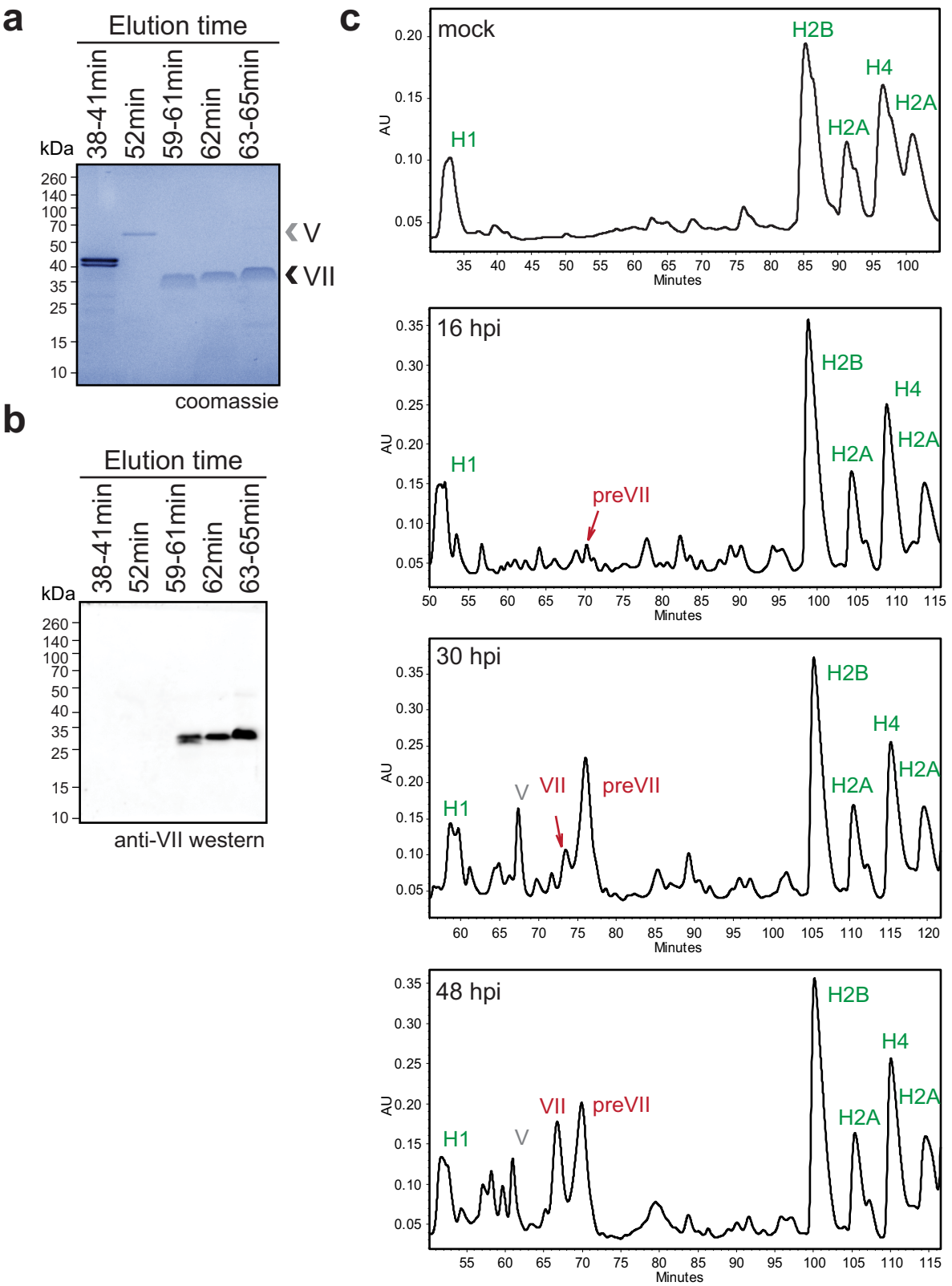
Extended Data Figure 2 | Protein VII associates tightly with chromatin and binds DNA and nucleosomes *in vitro*. **a**, Western blot analysis showing protein VII in histone extracts from infected HeLa cells at 24 hpi. **b**, Chromatin fractionation of lysates from A549 cells that were uninfected (mock) or infected for 24 h with Ad5. Viral and cellular proteins were detected by western blotting with various antibodies as indicated. **c**, Agarose gel analysis of DNA extracted from nuclear fractionation experiments, indicating that the size of DNA is between 100 and 200 bp and elutes predominantly in the higher-salt fractions. **d**, Chromatin fractionation of cells induced to express protein VII, indicating that protein VII is present in the highest-salt fraction from the first day of induction. **e**, **f**, Recombinant protein-VII–His binds DNA. Incubating increasing molar amounts of protein VII with 195 bp DNA results in shifts by native gel electrophoresis, indicating protein-VII–DNA complex formation. Staining with either ethidium bromide (**e**) or Coomassie (**f**) are shown to verify the presence of DNA and protein, respectively. **g**, Ethidium bromide staining shows DNA content of nucleosome shifts from gel in Fig. 1f.

© 2016 Macmillan Publishers Limited. All rights reserved



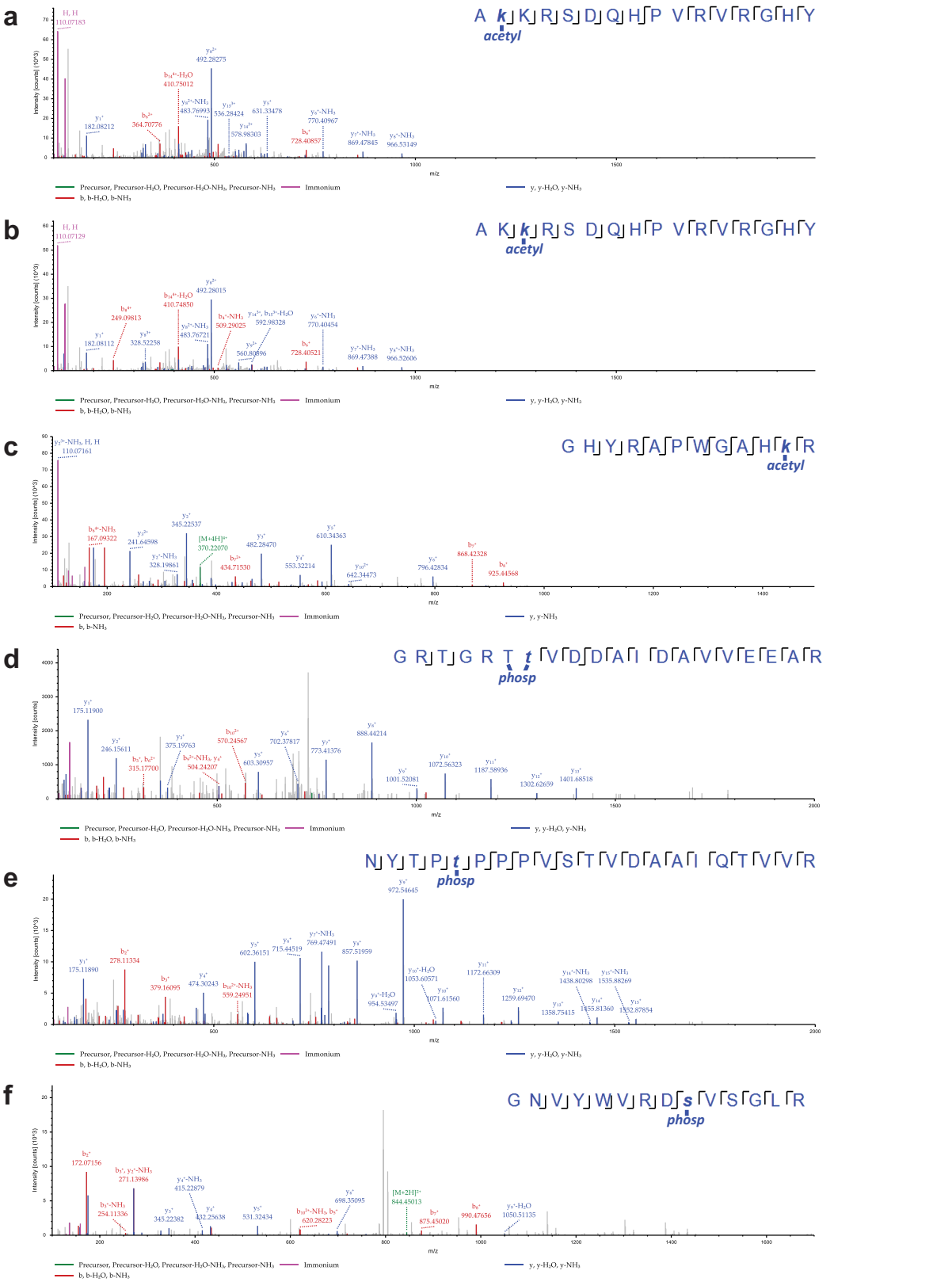
Extended Data Figure 3 | Bioanalyzer examination of MNase-digested nucleosomes and protein-VII–nucleosome complexes. a, 195 bp nucleosomes or protein-VII–nucleosome complexes were incubated with MNase for the indicated times, the reaction was stopped, DNA was extracted and analysed. As in Fig. 1g, nucleosomes are shown in black and protein-VII–nucleosome complexes in orange. The presence of protein VII pauses digestion at 165 bp, suggesting that protein VII is blocking access to the DNA. **b**, 147 bp nucleosomes or protein-VII–nucleosome complexes were incubated with MNase for the indicated times, the reaction was stopped, DNA was extracted and analysed. Graphs show nucleosomes in grey and protein-VII–nucleosome complexes in orange. The presence of protein VII completely blocks digestion even after nucleosomes alone have been digested well beyond the core particle. In contrast to what would be expected for linker histones, protein VII protects the core nucleosome particle from digestion. These data indicate that protein VII may be masking the substrate for MNase through complex formation. This represents a unique mechanism of nucleosome binding and suggests a model for blocking DNA access in cellular chromatin during infection.

© 2016 Macmillan Publishers Limited. All rights reserved



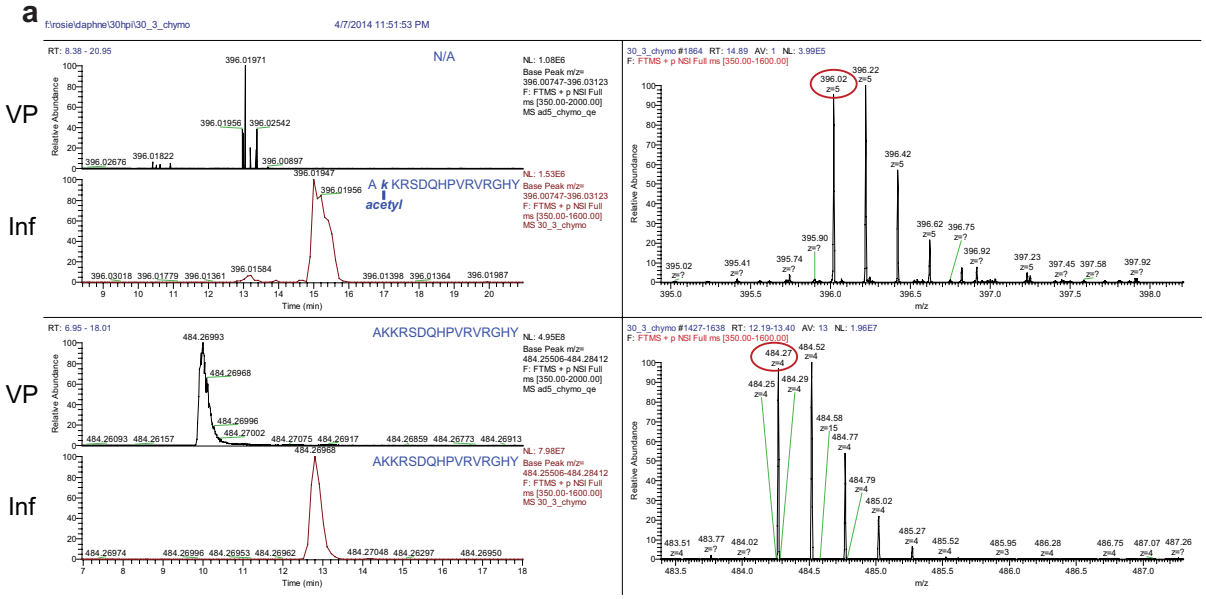
Extended Data Figure 4 | Purification of protein VII from infected cells. **a**, Coomassie-stained SDS–PAGE analysis of fractions from RP-HPLC in Fig. 2a. The bands in fraction 38–41 min correspond to histone H1. Protein VII and V, as indicated, were verified by MS analysis (data not shown). The slight upward shift of the protein VII bands in the later peak corresponds to the higher abundance of protein preVII, as seen by HPLC in Fig. 2a. **b**, Western blot analysis of protein VII in HPLC fractions from **a**. **c**, Time course of infection followed by histone extraction and HPLC analysis. MS analysis verified peaks in each sample as indicated.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | Representative mass spectra. a–f, Annotated MS/MS spectra of identified peptides of protein VII containing PTMs (**a–c**, acetylated peptides; **d–f**, phosphorylated peptides). The images represent the observed fragment ions collected using MS/MS collision-induced dissociation (CID). Coloured lines represent matches between observed and expected fragment ions of the given peptides. Specifically, green lines represent not fragmented precursor mass, blue lines represent matches with y-type fragments, red lines with b-type fragments, and yellow boxed masses represent fragments containing PTM neutral losses (for example, ions that lost the phosphorylation during fragmentation).

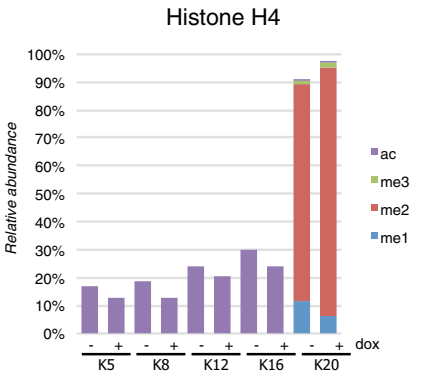
© 2016 Macmillan Publishers Limited. All rights reserved

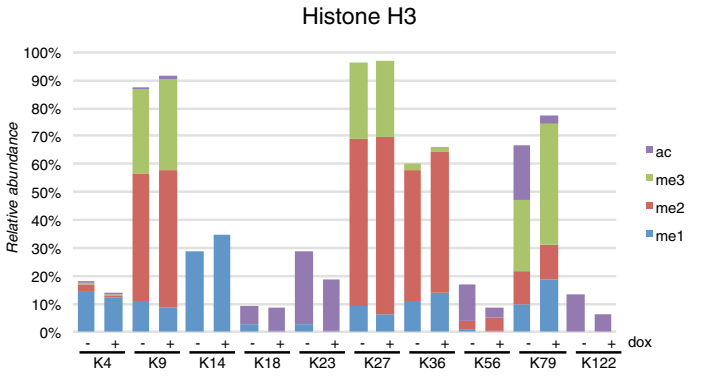


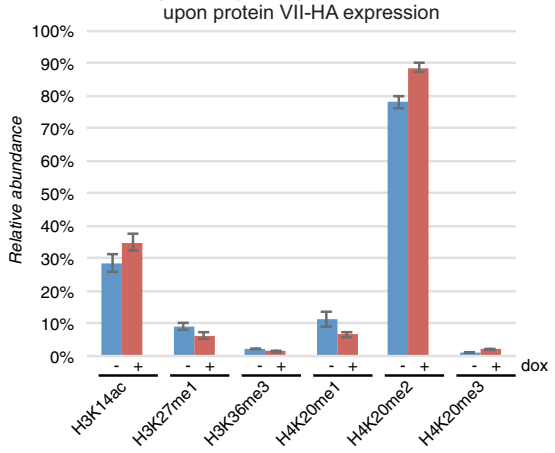
b

Post translational modifications of protein VII

Protein Name	Modification	Modification Site	Modified Sequence
Major core protein VII Accession Number 130683			
	Acetylation	2	A*k*KRSDQHPVRVRGHY
	Acetylation	3	AK*k*RSDQHPVRVRGHY
	Acetylation	24	GHYRAPWGAH*k*R
	Phosphorylation	31	GRTGR*t*TVDDAIDAVVEEAR
	Phosphorylation	48	NY*t*PTPPPVSTVDAAIQTVVRGARR
	Phosphorylation	50	NYTP*t*PPPVSTVDAAIQTVVR
	Phosphorylation	159	WVRD*s*VSGL
Major core protein VII precursor Accession Number 130748			
	Acetylation	19	FPS*k*MFGGAKKRSDQHPVR
	Phosphorylation	18	FP*s*KMFGGAKKRSDQHPVR
	Acetylation	25 (2)	GGA*k*KRSDQHPVRVRGHY
	Acetylation	26 (3)	GGAK*k*RSDQHPVRVRGHY
	Acetylation	47 (24)	GHYRAPWGAH*k*R
	Phosphorylation	54 (31)	GRTGR*t*TVDDAIDAVVEEAR
	Phosphorylation	71 (48)	NY*t*PTPPPVSTVDAAIQTVVR
	Phosphorylation	73 (50)	NYTP*t*PPPVSTVDAAIQTVVR
	Phosphorylation	182 (159)	WVRD*s*VSGL







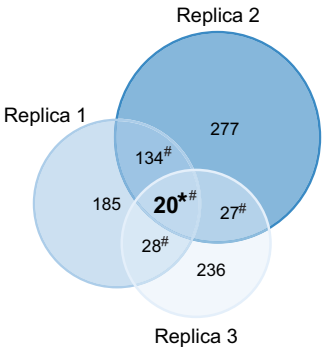
Extended Data Figure 6 | See next page for caption.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Figure 6 | Acetylated protein VII spectra from virus particles and analysis of total histone PTM changes upon protein VII expression. a, Liquid chromatography-mass spectrometry (LC-MS) analysis of unmodified and modified chymotryptic peptide AKKRSDQHPVRVRGHY. On the left, nano-LC-MS-extracted ion chromatograms of protein VII peptides identified in the histone extracts of adenovirus infected cells (Inf) or viral particles (VP). The top left represents the modified form, while the bottom left represents the unmodified form. Non-modified forms were detected in both conditions for Inf and VP, while the acetylated form was unique for the infected sample only (Inf). On the right, full MS spectrum of the modified (top) and unmodified (bottom) peptide. Circled mass represents the monoisotopic signal of the peptide. **b**, Summary of post-translational modifications detected on protein VII. Peptides shown were identified during infection at various time points with the mature protein VII in the top row and preVII in the bottom row. The numbers in brackets for preVII indicate the location of the same moiety in mature protein VII. Acetylation sites were detected in approximately 3% of peptides for mature protein VII and 2% of peptides in preVII. Phosphorylation was detected in approximately 1% of peptides for mature protein VII and preVII. **c**, **d**, Quantification of histone H3 (**c**) and H4 (**d**) PTMs in protein-VII–HA-induced (+dox) and -uninduced (−dox) A549 cells from the analysis of crude histone mixtures ($n = 3$ biological replicates). Positions of PTMs are listed along the x axis. Modification type is indicated by colour as shown. y Axis represents the cumulative extent of PTMs relative to the total histone H3 or H4, respectively. **e**, Breakdown of the histone marks (H3K14ac, H3K27me1, H3K36me3, H4K20me1, H4K20me2 and H4K20me3) found to be significantly different ($n = 3$ biological replicates) in terms of relative abundance between the protein-VII–HA-induced and -uninduced states (<5% homoscedastic two-tailed t-test). Mean ± s.d.

© 2016 Macmillan Publishers Limited. All rights reserved

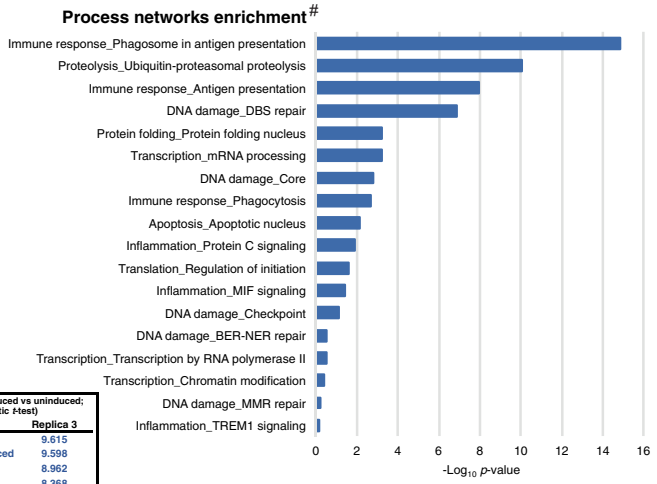
a



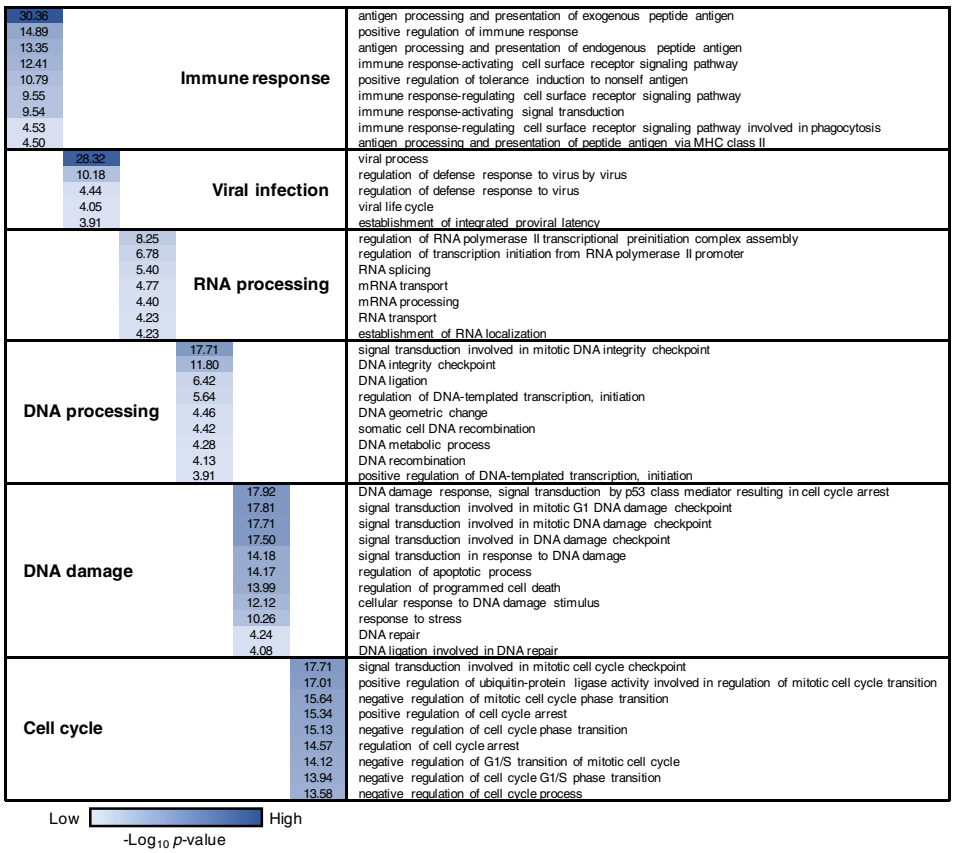
b

Protein*	Gene	Log2 fold change (VII-HA induced vs uninduced; <5% homoscedastic t-test) Replica 1	Replica 2	Replica 3
High mobility group protein B1	HMGB1	9.152	10.176	9.615
High mobility group protein B2	HMGB2	8.420	VII-HA induced	9.598
Homo sapiens Protein SET	SET	10.077	8.478	8.962
High mobility group protein B3	HMGB3	VII-HA induced	9.817	8.368
HLA class I histocompatibility antigen, A-34 alpha chain	HLA-A	1.165	1.386	5.297
Nucleosome assembly protein 1-like 1	NAP1L1	3.016	3.041	2.008
Nexilin	NEXN	1.348	0.621	1.467
Coiled-coil-helix-coiled-coil-helix domain-containing protein 2, mitochondrial	CHCHD2	1.419	1.991	1.256
Heterogeneous nuclear ribonucleoprotein F	HNRPF	0.783	0.449	1.118
Cofilin-1	HEL-S-15	1.352	1.529	1.029
Golgi integral membrane protein 4	GOLIM4	3.138	0.591	1.028
LIM domain and actin-binding protein 1	LIMA1	1.242	0.659	0.978
Heat shock 70 kDa protein 1A/1B	HEL-S-103	0.410	0.594	0.926
Vimentin	HEL113	0.911	0.296	0.736
Sideroflexin-3	SFXN3	0.271	0.205	0.684
Ras-related protein Ral-A	RALA	0.428	0.811	0.645
Signal recognition particle 9 kDa protein	SRP9	0.339	0.172	0.643
NAD(P) transhydrogenase, mitochondrial	NNT	0.289	0.444	0.562
Casein kinase II subunit beta	CSNK2B	0.852	1.089	0.543
Putative oxidoreductase GLYR1	N-PAC	0.591	0.189	0.428

c

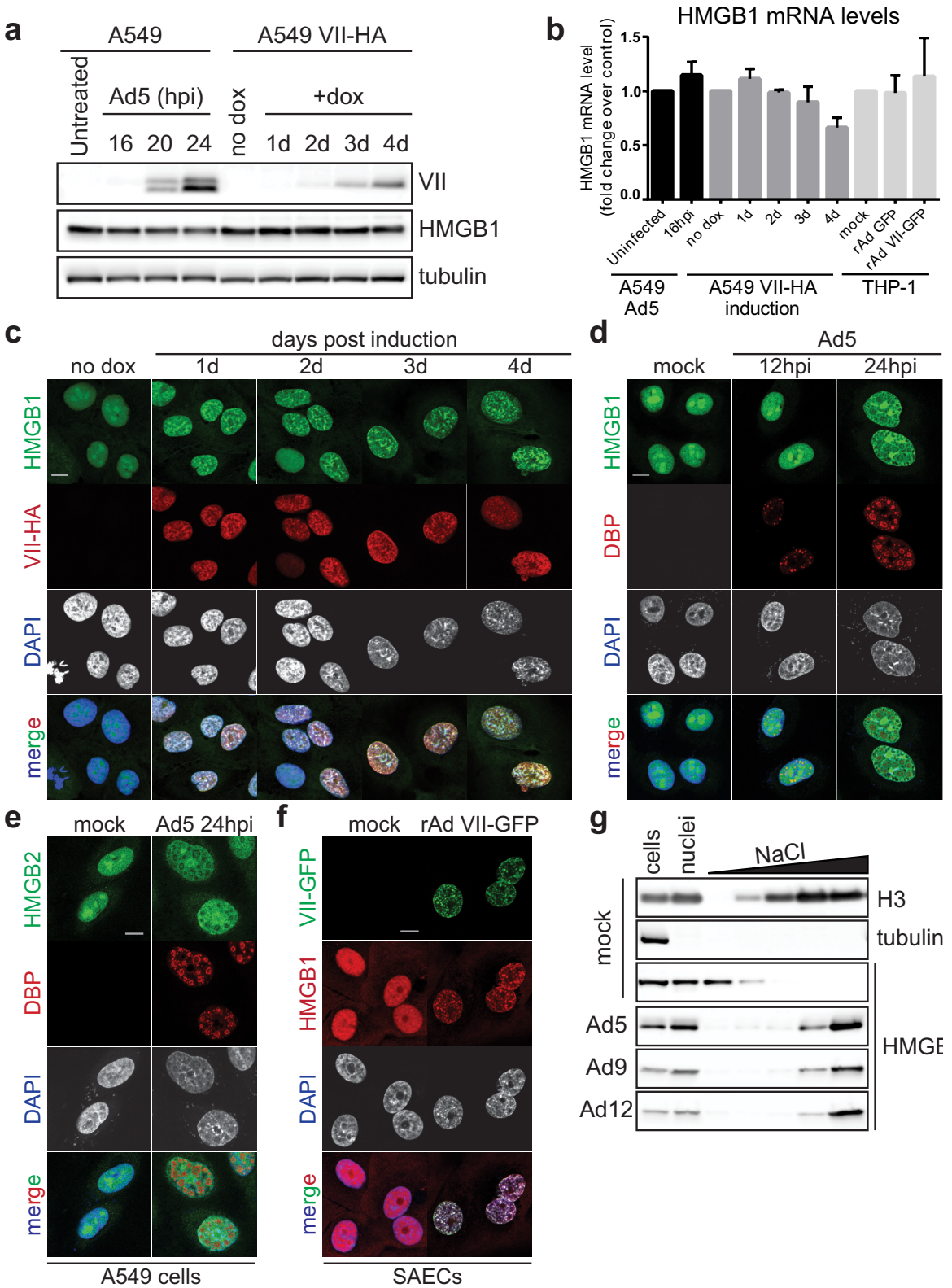


d



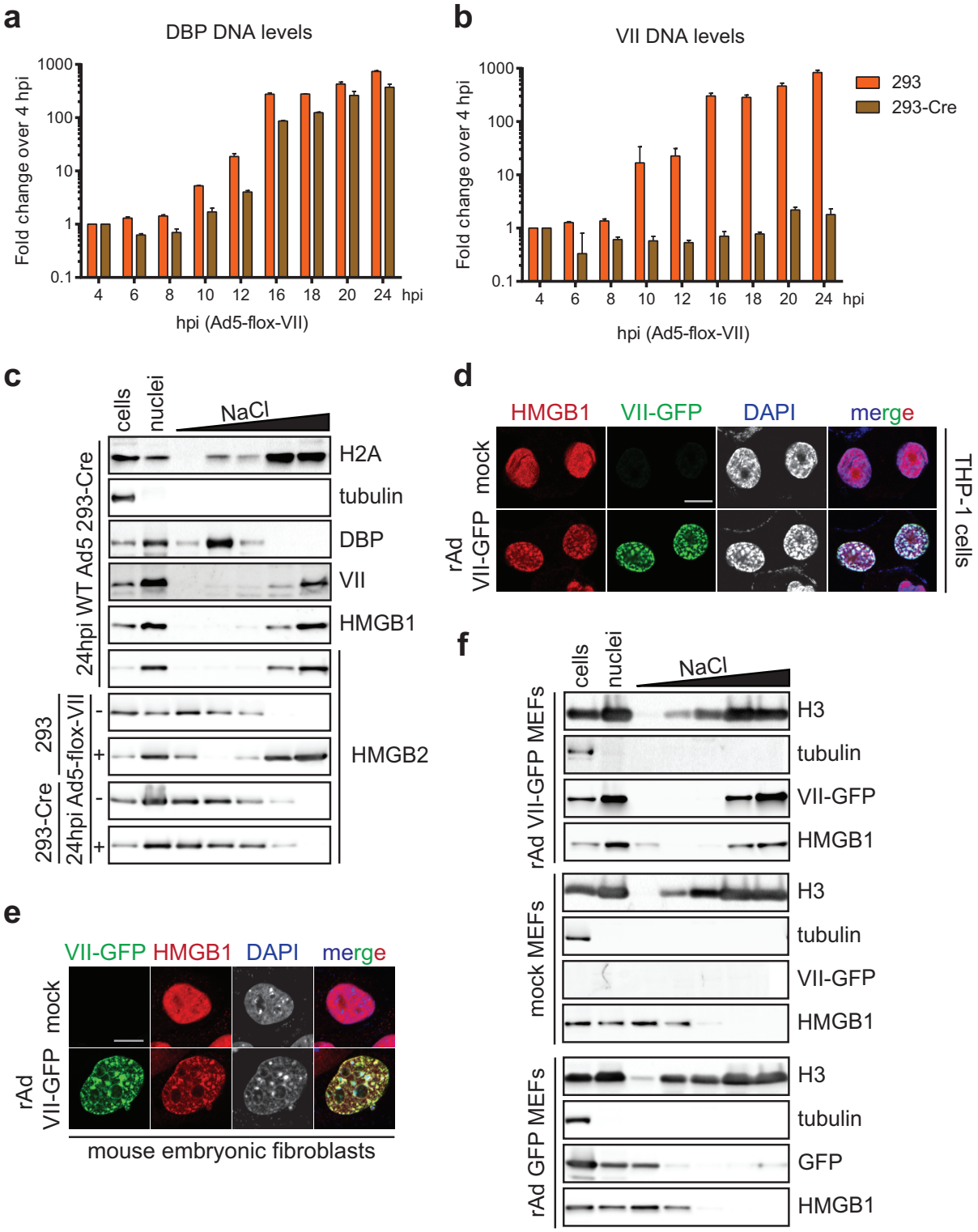
Extended Data Figure 7 | Bioinformatic analysis of proteins enriched in the high-salt fraction upon protein VII expression. a, Venn diagram showing overlap between three biological replicates of high-salt-fraction proteins significantly enriched compared with uninduced cells. **b**, Proteins found significantly enriched in the protein-VII–HA-induced state compared with uninduced (<5% homoscedastic *t*-test) in all three biological replicates ('VII–HA induced' indicates proteins identified only in protein-VII–HA-induced condition). **c**, **d**, Classification of proteins significantly enriched in minimum two out of three biological replicates (protein-VII–HA-induced versus uninduced) according to process network enrichment and Gene Ontology biological process (GeneGo MetaCore pathways analysis package; false discovery rate (FDR) < 5%); each Gene Ontology term was ranked using *P*-value enrichment.

© 2016 Macmillan Publishers Limited. All rights reserved



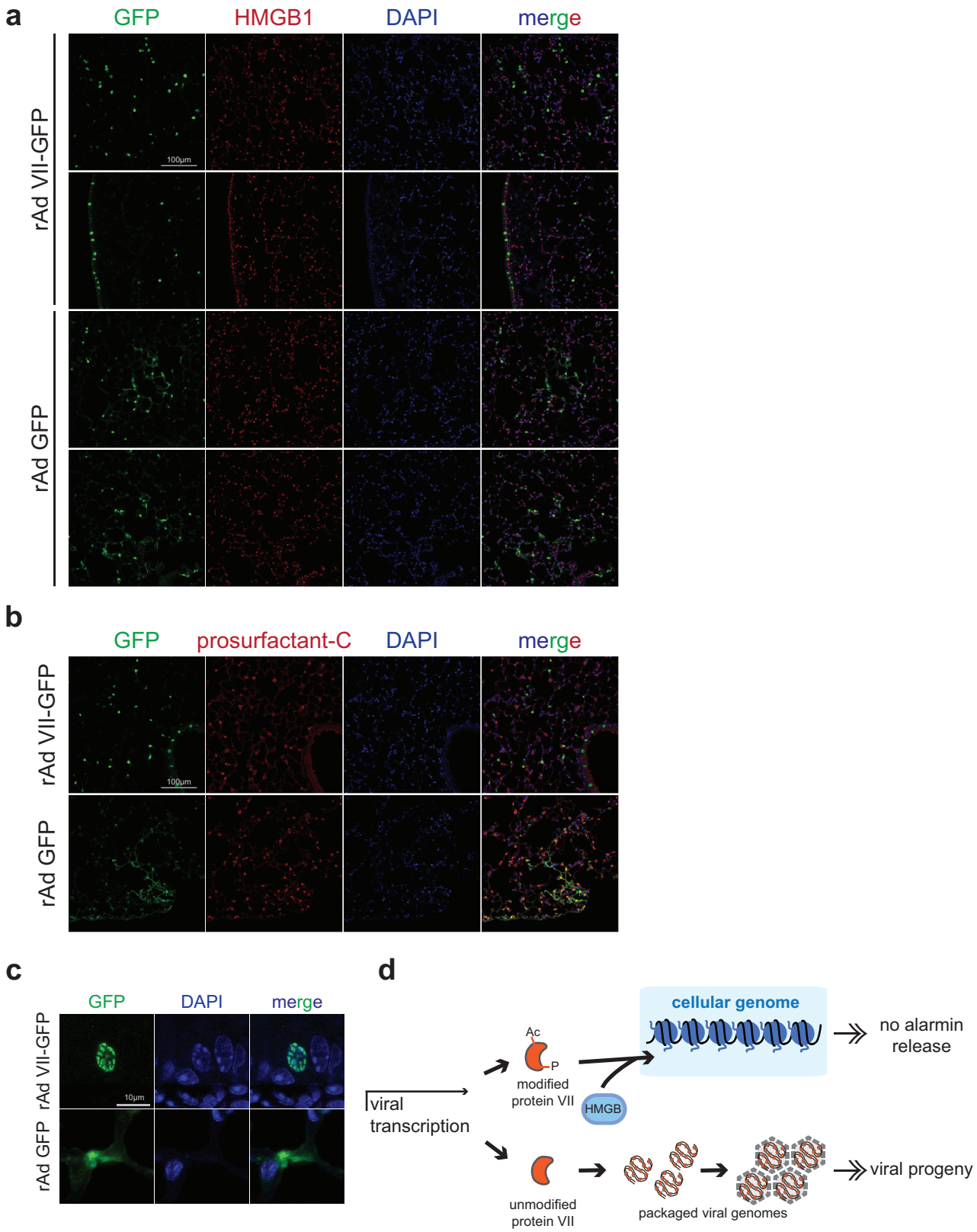
Extended Data Figure 8 | Protein VII retains HMGB1 and HMGB2 in chromatin. a, Western blot of adenovirus-infected or doxycycline-treated A549 cells showing the relative levels of protein VII expression. HMGB1 levels do not change upon infection or protein VII expression. Tubulin is shown as a loading control. **b**, Quantitative PCR analysis of mRNA transcripts of *HMGB1* in various cell types as indicated (for A549, $n = 3$ biological replicates; for THP-1, $n = 2$ biological replicates; mean ± s.d.). The levels of HMGB1 do not significantly change. **c**, Immunofluorescence analysis of a time course of protein-VII–HA (red) induction shown with HMGB1 (green) and DAPI (grey, blue in merge) in A549 cells. Expression of protein-VII–HA results in a change to the HMGB1 distribution upon expression. **d**, HMGB1 (green) localization changes between 12 and 24 hpi of wild-type adenovirus in A549 cells, and adopts a pattern similar to protein VII as in Fig. 1a. DBP (red) is shown as a marker of infection, DNA is stained with DAPI (blue in merge). **e**, Same as **d** showing that HMGB2 adopts the same pattern as HMGB1 during Ad5 infection at 24 hpi. **f**, Multiple cells showing the same pattern of HMGB1 relocalization upon expressing protein-VII–GFP as in Fig. 3g. **g**, HMGB1 retention in the high-salt fraction is conserved across adenovirus serotypes. Western blot analysis of HMGB1 from salt-fractionated A549 cells infected with Ad5, Ad9 or Ad12 as shown. Scale bars, 10 μm.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | Protein VII is necessary and sufficient for chromatin retention of HMGB1 in human and mouse cells. **a**, **b**, Replication of Ad5-flox-VII virus on 293 or 293-Cre cells. Quantitative PCR analysis of viral genomic DNA over a time course of infection (**a**) shows the DBP gene is increasing exponentially in 293 and 293-Cre cells when infected with Ad5-flox-VII virus. In contrast, PCR for the protein VII gene (**b**) demonstrates deletion in 293-Cre cells ($n = 2$ biological replicates, mean ± s.d.). **c**, Salt fractionation of 293-Cre cells infected with wild-type Ad5, indicating that the Cre recombinase does not interfere with the ability of protein VII to retain HMGB1 in the high-salt chromatin fraction. Protein VII is also necessary for the chromatin retention of HMGB2. **d**, THP-1 cells transduced to express protein-VII–GFP results in chromatin distortion and HMGB1 retention in chromatin. Immunofluorescence of transduced PMA-treated THP-1 cells showing protein-VII–GFP (green), HMGB1 (red) and DNA (grey, blue in merge). **e**, Transduction to express protein-VII–GFP is sufficient to relocalize mouse HMGB1 in mouse embryonic fibroblast (MEF) cells. **f**, Salt fractionation of mouse embryonic fibroblast cells transduced to express protein-VII–GFP. Human Ad5 protein VII is sufficient to retain mouse HMGB1 in the high-salt fraction in MEF cells. The control vector expressing GFP alone does not have this effect.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 10 | Transduction of mouse lungs demonstrating expression of GFP or protein-VII–GFP. a, Sections of mouse lungs transduced to express protein-VII–GFP or GFP co-stained for HMGB1. GFP signal shows multiple cell types transduced in both cases. Protein-VII–GFP has a more distinct nuclear signal than GFP, which also appears cytoplasmic. Two sections for each condition are shown to indicate transduction efficiency. **b**, Same as **a** but co-stained for prosurfactant-C to mark type II pneumocytes. Some cells are positive for both, confirming that multiple cell types were transduced. **c**, Zoomed images of individual epithelial cells from mouse lungs showing the characteristic protein-VII–GFP pattern colocalizing with DAPI in the nucleus. GFP only is mostly cytoplasmic. **d**, Schematic summarizing function of protein VII during infection. Newly synthesized protein VII late during infection can be post-translationally modified and binds to HMGB1, sequestering it on the cellular chromatin and preventing its release. Unmodified protein VII is packaged in viral progeny.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18316

The nature of mutations induced by replication–transcription collisions

T. Sabari Sankar[1,2]*, Brigitta D. Wastuwidyaningtyas[2]*, Yuexin Dong[1], Sarah A. Lewis[2] & Jue D. Wang[1,2]

The DNA replication and transcription machineries share a common DNA template and thus can collide with each other co-directionally or head-on[1,2]. Replication–transcription collisions can cause replication fork arrest, premature transcription termination, DNA breaks, and recombination intermediates threatening genome integrity[1–10]. Collisions may also trigger mutations, which are major contributors to genetic disease and evolution[5,7,11]. However, the nature and mechanisms of collision-induced mutagenesis remain poorly understood. Here we reveal the genetic consequences of replication–transcription collisions in actively dividing bacteria to be two classes of mutations: duplications/deletions and base substitutions in promoters. Both signatures are highly deleterious but are distinct from the previously well-characterized base substitutions in the coding sequence. Duplications/deletions are probably caused by replication stalling events that are triggered by collisions; their distribution patterns are consistent with where the fork first encounters a transcription complex upon entering a transcription unit. Promoter substitutions result mostly from head-on collisions and frequently occur at a nucleotide that is conserved in promoters recognized by the major σ factor in bacteria. This substitution is generated via adenine deamination on the template strand in the promoter open complex, as a consequence of head-on replication perturbing transcription initiation. We conclude that replication–transcription collisions induce distinct mutation signatures by antagonizing replication and transcription, not only in coding sequences but also in gene regulatory elements.

Mutations cause genetic diseases and drive evolution by altering either the gene-coding sequence or the noncoding elements that control gene expression. A variety of mechanisms underlie mutagenesis: DNA replication errors, error-prone repair, transcription-associated mutagenesis (TAM), and replication-stalling-mediated template switch[10,12–15]. Many mutagenic mechanisms depend on two fundamental processes—replication or transcription. However, little is known about the mutagenic mechanisms involving replication–transcription collision, an unavoidable outcome of the two processes sharing the same DNA template. Identifying the mutagenic consequences of replication–transcription collisions remains an important challenge owing to the difficulty of differentiating collision-induced mutation events from those induced by either replication or transcription.

An experimental approach to identifying collision-induced mutagenesis is to analyse the mutagenic consequence of altering the relative directionality of transcription to replication[7,10–12]. Head-on collisions are proposed to generate mutations more frequently than co-directional collisions, which may underlie the genome-wide bias for essential genes to be transcribed co-directionally to replication[5,7,11]. In support of this hypothesis, in the bacterium *Bacillus subtilis*—in which 94% of essential genes are co-directional[16]—base substitution rates are higher within genes oriented head-on than co-directional to replication[7,11]. However, the orientation-dependent difference in substitution rates can also be explained by the difference in the fidelity between leading- and lagging-strand replication[12,15,17,18], challenging the notion that collisions generate base substitutions in coding sequence[11,19]. Thus, conclusive evidence for collision-induced mutations is still lacking and necessitates a systematic analysis of collision-generated mutation signatures beyond base substitutions in the coding sequence.

Here, we investigate whether mutations are generated by collisions by identifying the signatures and characterizing the mechanisms of mutations caused by co-directional versus head-on collisions. We first developed an assay that can detect a wide range of mutations in *B. subtilis*. We chose the thymidylate synthetase gene *thyP3* because any complete loss-of-function mutation in *thyP3* can be selected using trimethoprim resistance (Extended Data Fig. 1a). To evaluate the effect of gene directionality on mutagenesis, we placed *thyP3* under an isopropyl-β-D-thiogalactoside (IPTG)-inducible promoter at a single location on the chromosome in either co-directional or head-on orientation (Fig. 1a). To estimate mutation rates, we performed the Luria–Delbrück fluctuation test using multiple growth cultures, selected for *thyP3* mutants after growth, and statistically determined the rate of spontaneous mutations in *thyP3* (Fig. 1b)[20,21]. Two additional features of our assay allowed critical analyses of mutagenesis. First, we chose the non-native, phage-encoded *thyP3* as the target sequence and deleted *thyA*, the native homologue of *thyP3*. We avoided using a native gene to evaluate the impact of gene directionality on mutagenesis because evolution may have already eliminated potential mutation hotspots within a native gene in its original orientation. Second, we took advantage of the temperature sensitivity of a second endogenous gene, *thyB*, to ensure that mutants were not defective during growth, which could complicate the measurement of mutation rate. We grew cells at the permissive temperature (37 °C), during which the functional ThyB masks any competitive disadvantage of *thyP3* mutations (Extended Data Fig. 1b). Selection was done at a non-permissive temperature under which ThyB is inactivated and phenotypes associated with *thyP3* mutation would be exposed (Fig. 1b and Extended Data Fig. 1c). In the presence of *thyB*, mutants follow the Luria–Delbrück distribution (Fig. 1c), demonstrating that mutations arise with a constant rate per cell division before and not after selection[20,21]. Notably, the use of *thyB* was critical because without *thyB*, mutants followed the Poisson distribution instead of the expected Luria–Delbrück distribution (Fig. 1c), presumably due to the growth defect of *thyP3* mutants (Extended Data Fig. 1d).

Using this assay, we compared mutations resulting from co-directional- versus head-on-oriented *thyP3*. When induced by IPTG, transcription reaches similar levels from either co-directional or head-on *thyP3* (Extended Data Fig. 2a). A ~60% increase in total mutation rate in the head-on *thyP3* compared with co-directional *thyP3* was observed (Fig. 1d and Extended Data Fig. 2b–e). Next, we sequenced ~2,000 mutants and obtained ~400 distinct mutations (Extended Data Figs 3 and 4). Only fewer than a third of mutations observed in *thyP3* under induced transcription were base substitutions within the coding region.

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. [2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved

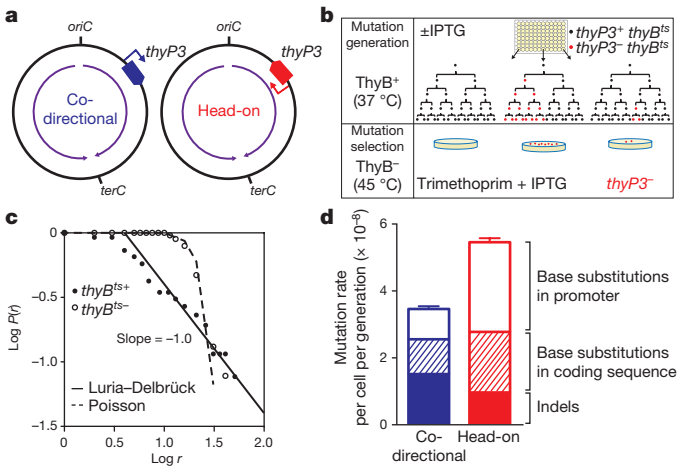


Figure 1 | Transcription directionality affects spontaneous mutation rates and spectra in *B. subtilis*. **a**, *thyP3* gene with an IPTG-inducible promoter is integrated into the chromosome either co-directionally or head-on to replication. Purple arrow indicates replication direction; *oriC* indicates replication origin; and *terC* indicates replication terminus. **b**, Modified fluctuation test to measure the rate of spontaneous mutations conferring trimethoprim resistance. *ts*, temperature sensitive. **c**, Distribution of mutants: number of mutants per culture (r) plotted against proportion of cultures with $\geq r$ mutants ($P(r)$). **d**, The mutation rates in co-directional and head-on *thyP3* (subdivided by mutation spectra) when transcription is induced with IPTG. Mutation rates are expressed as mean ± standard error of the mean (s.e.m.).

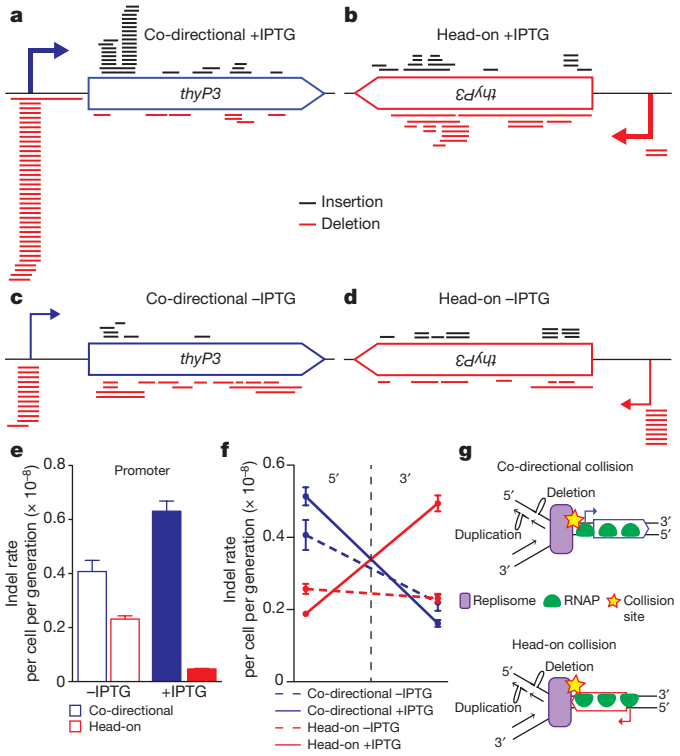


Figure 2 | Distribution of indels is strongly dependent on transcription directionality and strength. a–d, Positional distribution of indels of ≥3 bp in co-directional and head-on *thyP3* under induced (+IPTG) and uninduced (−IPTG) conditions. Each bar represents an insertion (black) or deletion (red). **e**, The rates of ≥3 bp indels at the promoter. **f**, The rates of ≥3 bp indels within 5′ (1–420 bp) and 3′ (421–840 bp) of the coding region. Rates of insertions are plotted separately from deletions in Extended Data Fig. 6d. **g**, Model illustrating the mechanism of generation of indels in the vicinity of the collision site in co-directional and head-on orientations, via fork slippage (shown here), template switch or fork reversal (Extended Data Fig. 6e). Error bars indicate s.e.m.

The remaining majority of mutations fall into two prominent classes: insertions/deletions (indels) and promoter base substitutions. Their mutation rates are strongly and differentially altered by transcription directionality and strength (Extended Data Fig. 2e). These alterations are mostly not due to competitive or selection bias of the mutants (Extended Data Fig. 5). Further analyses, described later, revealed that indels and promoter substitutions are probably induced by replication–transcription collisions.

Indels are probably generated upon stalling of a replication fork after collision with a transcription complex or a transcription factor[3,22]. First, the majority of indels are duplications/deletions between repeated DNA sequences (3–522 base pairs (bp); Extended Data Fig. 6a), which were proposed to originate from slippage or template switch of stalled replication forks[13,14]. Second, the frequencies of indels at different locations within *thyP3* are strongly influenced by its transcription orientation and strength (Fig. 2a–f and Extended Data Fig. 6b). When *thyP3* was co-directional to replication, indels were predominantly enriched at the promoter and 5′ half of the coding region (Fig. 2a), including promoter-proximal regions where RNA polymerases (RNAPs) are known to often pause[8]. In contrast, when *thyP3* was head-on, indels were found predominantly within the 3′ half (Fig. 2b), a bias that is largely absent when transcription was uninduced (at basal level) (Fig. 2c, d). This transcription-dependent enrichment pattern reflects the vicinity in which the replication fork first encounters a transcription complex upon entering a transcription unit (Fig. 2g and Extended Data Fig. 6c). Promoter deletions depended on the recombination protein RecA, and thus are mostly caused by recombination[13] after replication fork collision with the transcription initiation complex[22] or repressors[23] (Extended Data Fig. 7). However, the distribution of indels within the transcribed sequence was not affected by RecA, suggesting that recombination is not necessary for their generation. Instead, collision with the transcription elongation complex[3,8,24] stalls replication fork progression, which can induce fork slippage, template switch or fork reversal that leads to duplications/deletions, or by collision-generated DNA breaks[6–9] followed by microhomology-mediated break-induced replication or microhomology-mediated end joining (Extended Data Fig. 6e, f)[14]. Our work thus reveals the strong contribution of replication–transcription conflicts to the generation of indels.

We next analysed base substitutions in the coding sequence, which have been proposed to be generated by replication–transcription conflicts[11]. In contrast to indels, base substitutions within the coding sequence were not enriched near locations of replication–transcription collisions (Fig. 3a–d). In addition, base substitution rates were similar under induced transcription compared with basal levels when considering identical mutation target sites (Extended Data Fig. 8a). We again observed higher substitution rates in the coding sequence of head-on than co-directional genes[7,11], which are probably due to different replication fidelity between leading and lagging strands[12,18], although collisions cannot be ruled out as a source of these mutations.

By contrast with coding sequence substitutions, promoter base substitutions were elevated upon induction of transcription, suggesting that transcription initiation causes genome instability at the promoter (Fig. 3e and Extended Data Fig. 8b). Most strikingly, this increase in promoter substitution rates is much stronger (400%) for head-on than co-directional transcription, strongly suggesting that head-on collisions generate promoter substitutions. To examine the generality of this observation, we performed a genome-wide phylogenetic analysis to estimate nucleotide substitutions per site in promoters from multiple natural isolates of *Bacillus*. The analysis showed that promoters of head-on genes have higher nucleotide substitutions than promoters of co-directional genes (Fig. 3f). Thus, head-on transcription not only increases the mutation rate of the *thyP3* promoter sequence, but increases the mutation rate on a genome-wide scale in natural populations.

© 2016 Macmillan Publishers Limited. All rights reserved

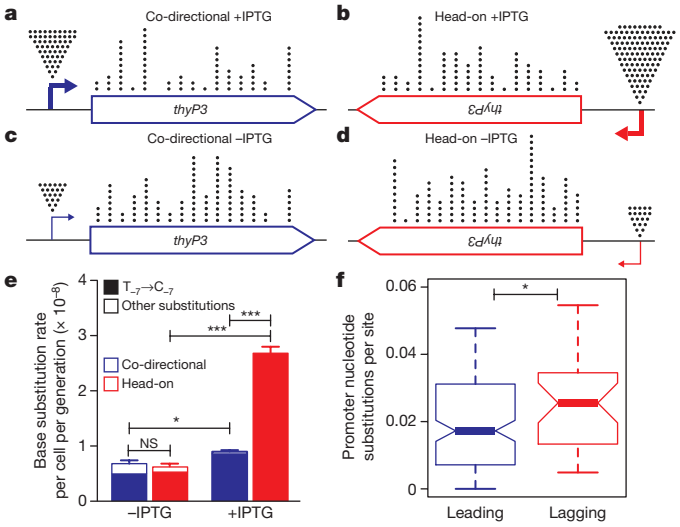


Figure 3 | Head-on transcription induces base substitutions at the promoter. a–d, Positional distribution of base substitutions in co-directional and head-on *thyP3* under induced (+IPTG) and uninduced (−IPTG) conditions. Each dot records a base substitution mapped in a 50-nucleotide window. **e**, The promoter base substitution rate is strongly increased in the head-on orientation upon IPTG induction. **f**, Mean nucleotide substitutions per site for each promoter, estimated pairwise among natural *Bacillus* isolates. Distribution for lagging-strand promoters ($n = 32$) compared with leading-strand promoters ($n = 147$). Nucleotide substitutions are also compared between promoters bound and not bound by transcriptional repressors (Extended Data Fig. 8c). Central mark of box plot represents median, edges are 25th and 75th centiles, notches are 95% confidence interval of median, and whiskers represent extreme data points within range. NS, not significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.0001$; Student's *t*-test, Mann–Whitney *U*-test. Error bars indicate s.e.m.

The most frequent substitution within the *thyP3* promoter is at a conserved nucleotide in the −10 element recognized by the major σ factor, T_{-7} (Fig. 4a). T_{-7}→C_{-7} substitution accounted for all promoter substitutions and 50% of total mutation events upon induced transcription of head-on *thyP3* (Fig. 3e). This enrichment is not due to competitive advantage of the C_{-7} mutant over wild type or other *thyP3* mutants (Fig. 4b and Extended Data Fig. 5a–c), indicating that T_{-7}→C_{-7} is a bona fide mutation hotspot obtainable with our assay. Importantly, T_{-7} is conserved across species and occurs in the promoters of ~50–70% of essential genes in *B. subtilis* and *Escherichia coli*[25]. The possibility that these promoters are all susceptible to transcription-induced T→C mutagenesis implicates a previously unidentified, pervasive mechanism that can inactivate the transcription of many genes and result in loss of viability. Indeed, in *E. coli* T_{-7}→C_{-7} was observed as a mutation hotspot in the head-on orientation in a plasmid-based assay (Extended Data Fig. 8d)[26], and T→C was also observed in other positions of *cis*-regulatory elements beyond the −7 position[27], suggesting that base substitutions in gene-regulatory elements is a signature of head-on transcription in bacteria.

To examine the mechanism underlying this mutation, we used a restriction-enzyme-based assay that exclusively detects T_{-7}→C_{-7} (Extended Data Fig. 8e) in *thyP3* to test several alternatives. First, we found that the error-prone DNA polymerase PolIV, which was proposed to be responsible for collision-induced substitutions[19], is not a major contributor of this mutation (Fig. 4c). Second, T_{-7}→C_{-7} is not generated by error-prone recombination repair as it still occurs frequently in the absence of *recA* (Extended Data Fig. 7e). Third, we examined whether a commonly occurring G–T wobble mismatch, which is generated by the replicative DNA polymerase and efficiently corrected by mismatch repair[28], accounts for this mutation. Inactivating mismatch repair increased the mutation rate of *thyP3* by ~60-fold, similar to other mutation assays[18], and increased T→C substitutions at hotspots in the coding sequence by ~1,000-fold (Extended Data Fig. 8f, g). Notably, we did not find any T_{-7}→C_{-7} substitutions upon screening ~1,000 mismatch repair mutants, suggesting that T_{-7}→C_{-7} is not generated via G–T mismatch.

After ruling out these known models of mutagenesis, we propose a new model that explains the frequent T_{-7}→C_{-7} substitutions on the basis of the structure of the bacterial promoter open complex where the −10 element is single-stranded[25,29,30]. Specifically, during transcription initiation, T_{-7} on the non-template strand is buried in a σ-factor pocket, and its complementary base on the template strand (A_{-7}) is unpaired and vulnerable to spontaneous deamination to hypoxanthine[27] (Fig. 4d). Hypoxanthine can base pair with cytosine during replication, leading to the T_{-7}→C_{-7} mutation. This model is further supported by our data that treating cells with nitrous acid, an inducer of base deamination, leads to an increased frequency of T_{-7}→C_{-7} mutation, which is more pronounced in the hypoxanthine-DNA glycosylase mutant (Fig. 4e), supporting hypoxanthine as the premutagenic intermediate. The cellular adenine deaminase is not a major factor responsible for

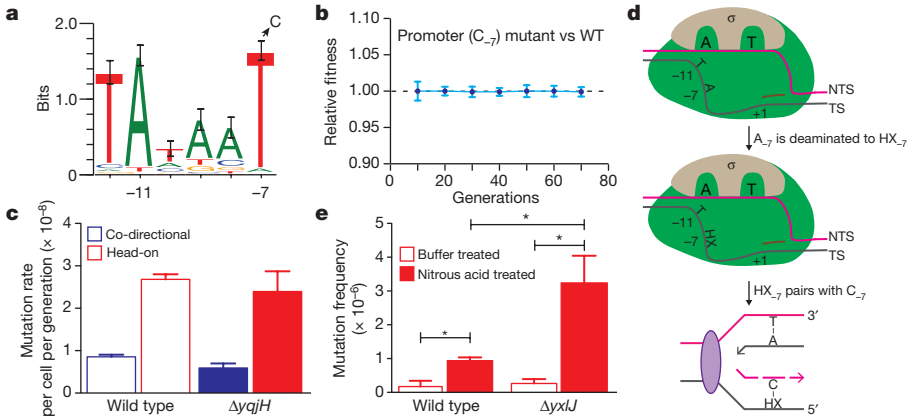


Figure 4 | Promoter T_{-7}→C_{-7} is a mutation hotspot generated via deamination. a, The consensus −10 element of *B. subtilis* SigA-dependent promoters ($n = 358$). The strongly conserved T_{-7} is frequently mutated to a C. **b**, Fitness of head-on T_{-7}→C_{-7} mutant relative to head-on wild-type (WT) *thyP3* cells under induced transcription (mean ± standard deviation (s.d.)). vs, versus. **c**, Mutation rate of T_{-7}→C_{-7} in *yqjH* mutant (error-prone polymerase PolIV). **d**, Model illustrating the mechanism of generation of T_{-7}→C_{-7}. During transcription initiation, the −10 element is single-stranded, T_{-7} on the non-template strand (NTS) is flipped into the σ-binding pocket, creating solvent accessibility for A_{-7} on the template strand (TS), allowing it to be deaminated to hypoxanthine (HX). HX base pairs with C during replication, resulting in T_{-7}→C_{-7}. **e**, T_{-7}→C_{-7} frequencies in head-on *thyP3* upon nitrous acid treatment of wild-type and Δ*yxlJ* (hypoxanthine-DNA glycosylase) strains. $*P < 0.05$; Student's *t*-test. Error bars indicate s.e.m.

© 2016 Macmillan Publishers Limited. All rights reserved

T_{-7}→C_{-7} mutation (Extended Data Fig. 8h), indicating that A_{-7} is spontaneously deaminated while sequestered within the transcription initiation complex. It is likely that other bases within the promoter open complex can also be mutated via deamination, although those mutations do not completely abolish gene expression and so cannot be identified by our assay. Our work thus uncovers a mechanism of promoter mutagenesis that implies a greater susceptibility of promoters to mutation than has previously been realised.

Our proposed mechanism represents a novel mutagenesis pathway that is distinct from TAM[10], which introduces substitutions within the transcribed sequence via deamination on the non-template strand, while the template strand of the coding sequence is protected by base pairing with nascent RNA (that is, RNA–DNA hybrid). In contrast, the promoter is upstream of the transcription start site, and thus is not protected by RNA–DNA hybrid but remains vulnerable to deamination or other premutagenic DNA damage upon open complex formation (Fig. 4d). We propose a model that head-on replication interferes with RNA polymerase escape from the promoter, rendering the promoter open complex more susceptible to premutagenic DNA damage, subsequently leading to mutations.

Our work reveals two types of collision-induced mutations, indels and promoter substitutions, which are generated by distinct mechanistic pathways probably resulting from mutual antagonism between replication and transcription upon collision. Furthermore, our work supports the hypothesis that collision-induced mutagenesis contributes to the evolution of the strong co-directional bias of essential genes[5] and reveals that orientation-biased promoter mutations underlie this conserved aspect of genome organization. We suspect that these mutation signatures have important implications not only for understanding the fitness and evolution of bacteria but also across domains of life including humans. Indels can lead to copy number variation, an important cause of genetic diseases. Mutations in *cis*-regulatory elements lead to misregulation of gene expression, and *cis*-regulatory elements are found to be more susceptible to mutagenesis than coding regions in eukaryotic genomes[17]. Thus, harmonizing replication with transcription is a key factor in fitness and genome evolution across domains of life.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 March 2015; accepted 11 May 2016.
Published online 29 June 2016.

1. French, S. Consequences of replication fork movement through transcription units *in vivo*. *Science* **258,** 1362–1365 (1992).
2. Liu, B. & Alberts, B. M. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* **267,** 1131–1137 (1995).
3. Vilette, D., Ehrlich, S. D. & Michel, B. Transcription-induced deletions in *Escherichia coli* plasmids. *Mol. Microbiol.* **17,** 493–504 (1995).
4. Prado, F. & Aguilera, A. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J.* **24,** 1267–1276 (2005).
5. Mirkin, E. V. & Mirkin, S. M. Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell. Biol.* **25,** 888–895 (2005).
6. Pomerantz, R. T. & O'Donnell, M. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature* **456,** 762–766 (2008).
7. Srivatsan, A., Tehranchi, A., MacAlpine, D. M. & Wang, J. D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet.* **6,** e1000810 (2010).
8. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* **146,** 533–543 (2011).
9. Merrikh, H., Machón, C., Grainger, W. H., Grossman, A. D. & Soultanas, P. Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470,** 554–557 (2011).
10. Kim, N. & Jinks-Robertson, S. Transcription as a source of genome instability. *Nature Rev. Genet.* **13,** 204–214 (2012)
11. Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E. & Merrikh, H. Accelerated gene evolution through replication-transcription conflicts. *Nature* **495,** 512–515 (2013).
12. Fijalkowska, I. J., Jonczyk, P., Tkaczyk, M. M., Bialoskorska, M. & Schaaper, R. M. Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc. Natl Acad. Sci. USA* **95,** 10020–10025 (1998).
13. Bruand, C., Bidnenko, V. & Ehrlich, S. D. Replication mutations differentially enhance RecA-dependent and RecA-independent recombination between tandem repeats in *Bacillus subtilis*. *Mol. Microbiol.* **39,** 1248–1258 (2001).
14. Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nature Rev. Genet.* **10,** 551–564 (2009).
15. Kunkel, T. A. Evolving views of DNA replication (in)fidelity. *Cold Spring Harb. Symp. Quant. Biol.* **74,** 91–101 (2009).
16. Rocha, E. P. C. & Danchin, A. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nature Genet.* **34,** 377–378 (2003).
17. Reijns, M. A. M. *et al.* Lagging-strand replication shapes the mutational landscape of the genome. *Nature* **518,** 502–506 (2015).
18. Schroeder, J. W., Hirst, W. G., Szewczyk, G. A. & Simmons, L. A. The effect of local sequence context on mutational bias of genes encoded on the leading and lagging strands. *Curr. Biol.* **26,** 692–697 (2016).
19. Million-Weaver, S. *et al.* An underlying mechanism for the increased mutagenesis of lagging-strand genes in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **112,** E1096–E1105 (2015).
20. Luria, S. E. & Delbrück, M. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28,** 491–511 (1943).
21. Rosche, W. A. & Foster, P. L. Determining mutation rates in bacterial populations. *Methods* **20,** 4–17 (2000).
22. Mirkin, E. V., Castro Roa, D., Nudler, E. & Mirkin, S. M. Transcription regulatory elements are punctuation marks for DNA replication. *Proc. Natl Acad. Sci. USA* **103,** 7276–7281 (2006).
23. Vilette, D., Uzest, M., Ehrlich, S. D. & Michel, B. DNA transcription and repressor binding affect deletion formation in *Escherichia coli* plasmids. *EMBO J.* **11,** 3629–3634 (1992).
24. Tehranchi, A. K. *et al.* The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* **141,** 595–605 (2010).
25. Feklistov, A. & Darst, S. A. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* **147,** 1257–1269 (2011).
26. Yoshiyama, K., Higuchi, K., Matsumura, H. & Maki, H. Directionality of DNA replication fork movement strongly affects the generation of spontaneous mutations in *Escherichia coli*. *J. Mol. Biol.* **307,** 1195–1206 (2001).
27. Schaaper, R. M., Danforth, B. N. & Glickman, B. W. Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. *J. Mol. Biol.* **189,** 273–284 (1986).
28. Lu, A. L., Clark, S. & Modrich, P. Methyl-directed repair of DNA base-pair mismatches *in vitro*. *Proc. Natl Acad. Sci. USA* **80,** 4639–4643 (1983).
29. Zhang, Y. *et al.* Structural basis of transcription initiation. *Science* **338,** 1076–1080 (2012).
30. Zuo, Y. & Steitz, T. A. Crystal structures of the *E. coli* transcription initiation complexes with a complete bubble. *Mol. Cell* **58,** 534–540 (2015).

Acknowledgements We thank E. Robleto, R. Yasbin and L. Simmons for strains, M. Cox, R. Gourse, C. Hittinger, R. Landick, K. Wasserman, C. Gross, M. Laub, S. Rosenberg, L. Simmons and the Wang laboratory for discussions and comments on the manuscript. This work was supported by the National Institutes of Health Director's New Innovator Award DP2OD004433 to J.D.W.

Author Contributions J.D.W. conceptualized the study. T.S.S., B.D.W. and J.D.W. designed the experiments. T.S.S. performed *thyP3* fluctuation tests and sequencing of *recA*, *yqjH*, *mutSL* and *adeC* mutants, comparative genomic analyses, nitrous acid mutagenesis, competition assays and plating efficiency of mutants. B.D.W. developed the forward mutation assay, fluctuation tests and sequencing of wild-type strains, qRT–PCR, nalidixic acid fluctuation test, doubling time measurements and developed the restriction digest screening. Y.D. assisted the competition assay, plating efficiency and *mutSL* fluctuation tests. S.A.L. performed *thyP3* fluctuation tests with B.D.W. T.S.S., B.D.W. and J.D.W. analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.W. (wang@bact.wisc.edu).

Reviewer Information *Nature* thanks S. Mirkin, E. Nudler and R. Pomerantz for their contribution to the peer review of this work.

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Media and growth conditions. Unless otherwise indicated, cells were grown in S7 defined medium[31] containing 50 mM MOPS and supplemented with 1% glucose, 0.1% glutamate, 40 μg ml^{-1} tryptophan and 20 μg ml^{-1} thymine (Sigma-Aldrich) at 37 °C with vigorous shaking, and plated on solid medium (Spizizen's medium), supplemented with 1% glucose, 0.1% glutamate, 40 μg ml^{-1} tryptophan and 20 μg ml^{-1} thymine. Trimethoprim (RPI Research Products International) was added to plates at a final concentration of 5 μg ml^{-1} for selecting loss-of-function mutations in *thyP3* gene. To induce expression of *thyP3*, IPTG was added to the medium at a final concentration of 1 mM. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Strain construction. Strains used are derivatives of the wild-type strain *B. subtilis* 168 (JDW437) unless otherwise stated and are listed in Extended Data Table 1. The plasmids and PCR primers are listed in Extended Data Tables 1 and 2, respectively. The *thyP3* strains were created in the Δ*thyA* (JDW1543) background. *thyA* was deleted using the markerless deletion method[32] with plasmid pJW395. The head-on *thyP3* strain JDW1544 was generated by transforming JDW1543 with linearized plasmid pJW396. The co-directional *thyP3* strain JDW1563 was generated by transforming JDW1543 with linearized plasmid pJW397. Swapped head-on and co-directional *thyP3* strains JDW1900 and JDW1901 were created by transforming JDW1543 with linearized plasmid pJW430 and pJW431, respectively. The head-on *thyP3* strain (JDW1176) in a Δ*thyA* Δ*thyB* background was created by transforming JDW942 with linearized plasmid pJW331. The *lacZ* reporter strains used in competition assays were created by transforming the respective *thyP3* wild-type or mutant strains with linearized plasmid pJW417.

Plasmid pJW395 was constructed to create a markerless deletion of *thyA*, by inserting *thyA* upstream homologous sequence (PCR amplified by primers oJW1052/oJW1053) and downstream homologous sequence (PCR amplified by primers oJW1054/oJW1055) between the EcoRI and BamHI sites of pJW299. Plasmid pJW331 was constructed by inserting the *thyP3* gene between the SalI and SphI sites of pDR90. The *thyP3* gene sequence, including its promoter, was amplified from genomic DNA of JDW941 using oJW760/oJW761. Plasmid pJW396 was constructed by inserting the *thyP3* gene between the SalI and SphI sites of pDR110. Plasmid pJW397 was constructed by excising out the P_{spank}-*thyP3* region from pJW396 by double-restriction digest with EcoRI and SphI and replacing it with the P_{spank}-*thyP3* sequence in the inverse orientation between the EcoRI and SphI sites. The P_{spank}-*thyP3* sequence for inversion was amplified from pJW396 using primers oJW785 and oJW1137.

Plasmid pJW430 was created by Gibson assembly[33] of a DNA fragment containing the *lacI*-P_{spank}-*thyP3-spec* sequences and the portion of the pDR110 plasmid backbone containing the plasmid replication origin, amp^R, and *amyE* front (5′) and back (3′) homology sequences. The DNA fragment with *lacI*-P_{spank}-*thyP3-spec* sequences was amplified from pJW397 using oJW1336/oJW1339. The pDR110 backbone fragment was amplified from pDR110 using oJW1337/oJW1338. Plasmid pJW431 was created in the same way as plasmid pJW430, except the DNA fragment with *lacI*-P_{spank}-*thyP3-spec* sequences was amplified from pJW396, instead of pJW397, using the same primers. Plasmid pJW417 was created by Gibson assembly[33] of a DNA fragment containing the *spoVG-lacZ* sequences and a portion of pDR110 plasmid containing the P_{pen} promoter and *lacA* locus 5′ and 3′ homology sequences for integration. The DNA fragment containing the *spoVG-lacZ* sequence and plasmid backbone were amplified from pEX44 using oJW1200/oJW1201 and oJW1213/oJW1214, respectively. The P_{pen} promoter was amplified from pDR110 using oJW1202/oJW1203. The *lacA* 5′ and 3′ homology regions were amplified from the chromosomal DNA of *B. subtilis* 168 using oJW1215/oJW1199 and oJW1204/oJW1216, respectively.

Deletion mutants of *yqjH* gene encoding PolIV (JDW2266), *adeC* encoding adenine deaminase (JDW2501), *recA* encoding the recombinase RecA (JDW2288) and *yxlJ* encoding hypoxanthine-DNA glycosylase (JDW2284) were obtained from the *Bacillus* genetic stock centre (BGSC). Co-directional (JDW1563) and head-on *thyP3* (JDW1544) strains were transformed with the genomic DNA of each mutant and were selected on erythromycin plates at 37 °C. Deletion of each gene was confirmed by PCR (*yqjH*-oJW1900/1901; *adeC*-oJW1904/1905; *recA*-oJW2008/2009; *yxlJ*-oJW1906/1907) and *recA* mutant was also tested for ultraviolet sensitivity. Deletion of mismatch repair genes *mutS* and *mutL* was created by transforming the genomic DNA of JDW1297 into co-directional (JDW1563) and head-on *thyP3* (JDW1544) strains and were selected on kanamycin plates at 37 °C. The kanamycin gene insertion inactivated both mismatch repair genes and insertion was confirmed by PCR (oJW1902/1903).

Forward mutation fluctuation tests. Fluctuation tests were performed to measure the forward mutation rate. All the *thyP3* strains were in the background Δ*thyA thyB*$^+$. For each biological repeat, at least 30 parallel cultures of 0.1 ml in 96-well plates were set up for each strain at a dilution of 1×10^{-5} and grown at 37 °C to $OD_{600\,nm} = 0.4–0.6$ in S7 minimal medium with 20 μg ml^{-1} thymine and with 1 mM IPTG (induced transcription) or without IPTG (uninduced transcription). Loss-of-function mutations in *thyP3* genes confer resistance to trimethoprim (TMP). For selection of mutants, 0.1 ml of culture was plated on Spizizen's minimal medium containing 20 μg ml^{-1} thymine, 1 mM IPTG and 5 μg ml^{-1} trimethoprim. Plates were incubated at 45 °C, and the number of trimethoprim-resistant colonies were counted at 48 h (day 2) and 72 h (day 3) of incubation. Serial dilutions of at least three cultures were plated on non-selective medium to determine the average c.f.u. The number of mutations per culture (m) was estimated using the Ma–Sandri–Sarkar maximum likelihood estimator (MSS-MLE) method through the Fluctuation AnaLysis CalculatOR (FALCOR) web tool[34], and the mutation rate per cell per generation was calculated by $m/(2 \times N_t)$, where N_t is the average number of cells across cultures in a fluctuation test[21]. Fluctuation tests of the deletion mutants were performed as described for the wild-type strains earlier except for *recA* and *mutSL* deletion strains. Since *recA* mutant showed increased sensitivity to trimethoprim, selection of *thyP3* mutants was done at 1 μg ml^{-1} concentration of trimethoprim and mutant colonies were obtained from day 4 and day 5 after incubation. Fluctuation tests with *mutSL* mutants were performed identical to wild type, except that the cultures were diluted 1:20 for selection on trimethoprim plates, as inactivation of mismatch repair increases the mutation rate. The mean of mutation rates from $n \geq 3$ independent experiments was plotted with error bars representing standard error of the mean (s.e.m.). Statistical significance was calculated by paired Student's t-test of $\ln(m)$ values[21].

We employed a mutation assay for nalidixic acid resistance, which is conferred by mutations in the *gyrA* gene encoding DNA gyrase, to examine whether mutation rate is different outside the *thyP3* locus between the co-directional and head-on *thyP3* strains. For measurement of the mutation rate for nalidixic acid resistance (Nal^R), at least 30 parallel 1 ml cultures were grown in test tubes to $OD_{600\,nm} = 0.4–0.6$ and entire cultures were plated on minimal medium containing 20 μg ml^{-1} thymine, 1 mM IPTG and 50 μg ml^{-1} nalidixic acid (Sigma-Aldrich). Plates were incubated at 45 °C for 48 h, and the number of plates with no Nal^R colonies was counted. Serial dilutions were plated on non-selective medium to count the number of c.f.u. The number of Nal^R mutations per culture (m) was estimated using the P_0 method and the mutation rate was calculated by $m/(2 \times N_t)$[21]. Error bars represent the standard error from at least three independent experiments.

Mutation spectra and rates of different mutations. To obtain the mutation spectrum, genomic DNA from one colony per selective plate was extracted by using the *prep*GEM Bacteria kit (Zygem) and *thyP3* was PCR amplified and sequenced using primers oJW1013 and oJW1335. The rate of individual mutation was determined by multiplying the total mutation rate by the proportion of different mutations in the mutation spectra as described[35]. Statistical significance of differences between co-directional and head-on strains for different mutation types was obtained using Student's t-test.

qRT–PCR. Measurement of *thyP3* transcription levels was performed by qRT–PCR. Cultures were grown in minimal media with 20 μg ml^{-1} thymine, with or without 1 mM IPTG, to $OD_{600\,nm}$ 0.4–0.6. RNA was isolated using the Qiagen RNeasy kit and reverse-transcribed using SuperScript III reverse transcriptase (Life Technologies). Real-time PCR was performed using SYBR green master mix (Applied Biosystems) with primers oJW1217/oJW1218 for amplifying the beginning of *thyP3*. The *accA* gene transcript amplified with primers oJW1221/oJW1222 was used as an internal control[36].

Competition assay. Competition experiments were performed between strains carrying the wild-type and mutant *thyP3*. Strains were grown in S7 minimal medium supplemented with 1% glucose, 0.1% glutamate, 40 μg ml^{-1} tryptophan and 1 mM IPTG. Strains in competition were distinguished by integrating a *lacZ* reporter gene at the *lacA* locus in the chromosome, enabling the competitors to be distinguished on 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) indicator plates in which $LacZ^-$ and $LacZ^+$ form white and blue colonies, respectively. The *lacZ* marker was swapped between the competing strains to negate any growth effect from the *lacZ* marker. Strains were preconditioned in the growth medium to saturation. Cultures were then mixed in a 1:1 ratio, and serial passage was performed with 1:1,000 dilutions (~10 generations per cycle) every 12 h until 70 generations. The ratio of mutant over wild type at each cycle was estimated by plating the serially diluted cultures on SPII minimal plates supplemented with 40 μg ml^{-1} tryptophan, 20 μg ml^{-1} thymine and 40 μg ml^{-1} X-gal at 37 °C. Growth rate was calculated using the initial and final cell densities for each strain in the pair, and the relative fitness was calculated as the ratio of growth rates of mutant over wild-type cells[7]. Assays were performed with three independent replicates of the mutant tagged with *lacZ* and another three in which the wild type was tagged with *lacZ*. Relative fitness was then expressed as the mean ± s.d. of replicates with and without the marker. To rule out reversion of the mutant *thyP3* strain during competition growth, strains were plated at the end of 70 generations on X-gal indicator plates with and without thymine at 45 °C and also on trimethoprim plates;

© 2016 Macmillan Publishers Limited. All rights reserved

only the wild type formed colonies on plates without thymine and the mutants formed colonies only on plates with thymine. As expected, wild type did not form colonies on trimethoprim, while the mutants formed colonies.

Restriction digest screen of promoter mutation. To screen for $T_{-7}\rightarrow C_{-7}$ mutation in the promoter, the first half of the *thyP3* fragment including the promoter region was PCR amplified with primers oJW1335 and oJW1011 from mutant DNA. The PCR fragment was digested with AflIII enzyme (NEB) and digested products were analysed in 1.5% agarose gel. PCR fragments containing the $T_{-7}\rightarrow C_{-7}$ mutation are digested by AflIII, whereas wild-type fragments are not digested (Extended Data Fig. 8e).

Sequence logo of the −10 element of SigA-dependent promoters. We obtained the sequences of the −10 element of all experimentally validated SigA-dependent promoters ($n = 358$) available at the DBTBS database[37] and used the WebLogo tool[38] to generate the consensus motif with the default parameters to show the genome-wide conservation of the −10 element.

Comparative genomic and molecular evolutionary analyses. For comparative genomic and evolutionary analyses, we used the completed genomes of eight strains of *B. subtilis* and one *B. amyloliquefaciens* strain, a close relative of *B. subtilis*. The analysed genomes are listed in Extended Data Table 2. Complete genomes, amino acid and nucleotide sequences of genes and intergenic sequences, and gene annotation information were downloaded from the Integrated Microbial Genomes (IMG) database[39]. Core genes from *B. subtilis* and *B. amyloliquefaciens* were identified by standard all-against-all reciprocal best-hit method using BLASTP. Best bidirectional hits were considered when the alignment had >85% identity with 85% coverage length at an e-value cut-off of 10^{-20}. We eliminated any gene annotated as pseudogene and containing ambiguous nucleotides from the analysis.

To assign genes to leading and lagging strand, we obtained the sequence coordinates of *oriC* and *dif* sites from the DoriC database[40] for each genome, and using these coordinates in combination with transcript orientation information from the genome annotation files (plus or minus strand), genes were assigned to leading and lagging strands. All genes analysed were present on the same strand (either leading or lagging) in all the genomes analysed.

To extract promoter sequences, experimentally validated promoter annotations were obtained for the core genes of *B. subtilis* strain 168 from the DBTBS database[37]. Sequence encompassing the transcription start site (+1), the −10 and −35 elements of the promoter was obtained. Using these promoter sequences as references, homologous promoters from the other genomes of *B. subtilis* and *B. amyloliquefaciens* were obtained using the blastn-short algorithm of BLASTN employing the 75% identity over 80% alignment coverage with e-value less than 10^{-5}. We obtained 179 promoters (147 and 32 for leading- and lagging-strand genes, respectively).

The amino acid sequences and the corresponding nucleotide sequences of protein-coding core genes were aligned using the G-INS-i algorithm of the MAFFT alignment program (v.7.012b)[41]. Furthermore, to produce high-quality alignments, we used the PAL2NAL program (v.12.1)[42], which produces codon-based alignments from aligned protein sequences and the corresponding DNA sequences. Additionally, PAL2NAL reports whether the protein and nucleotide sequences have mismatches or in-frame stop codons. The codon-based alignments of the core genes generated by PAL2NAL did not contain any mismatches or in-frame stop codons, which ensured the high quality of the alignments.

For aligning the promoters, we used the E-INS-i algorithm of the MAFFT alignment program, which is optimized for aligning highly conserved motifs interspaced between weakly conserved regions. The alignments of the experimentally validated promoters were manually inspected for any misalignments.

Estimation of nucleotide substitutions in promoters. To estimate nucleotide substitutions in promoters, we first constructed phylogeny using the concatenated sequence of the core genome genes, that is, genes present in all the analysed genomes. The aligned nucleotide sequences were concatenated to create a single sequence for each analysed strain. Phylogeny was constructed using PhyML program[43] with 500 bootstrap replicates. The substitution model used was the general time reversible model (GTR) with discrete gamma model, and gamma parameter was estimated.

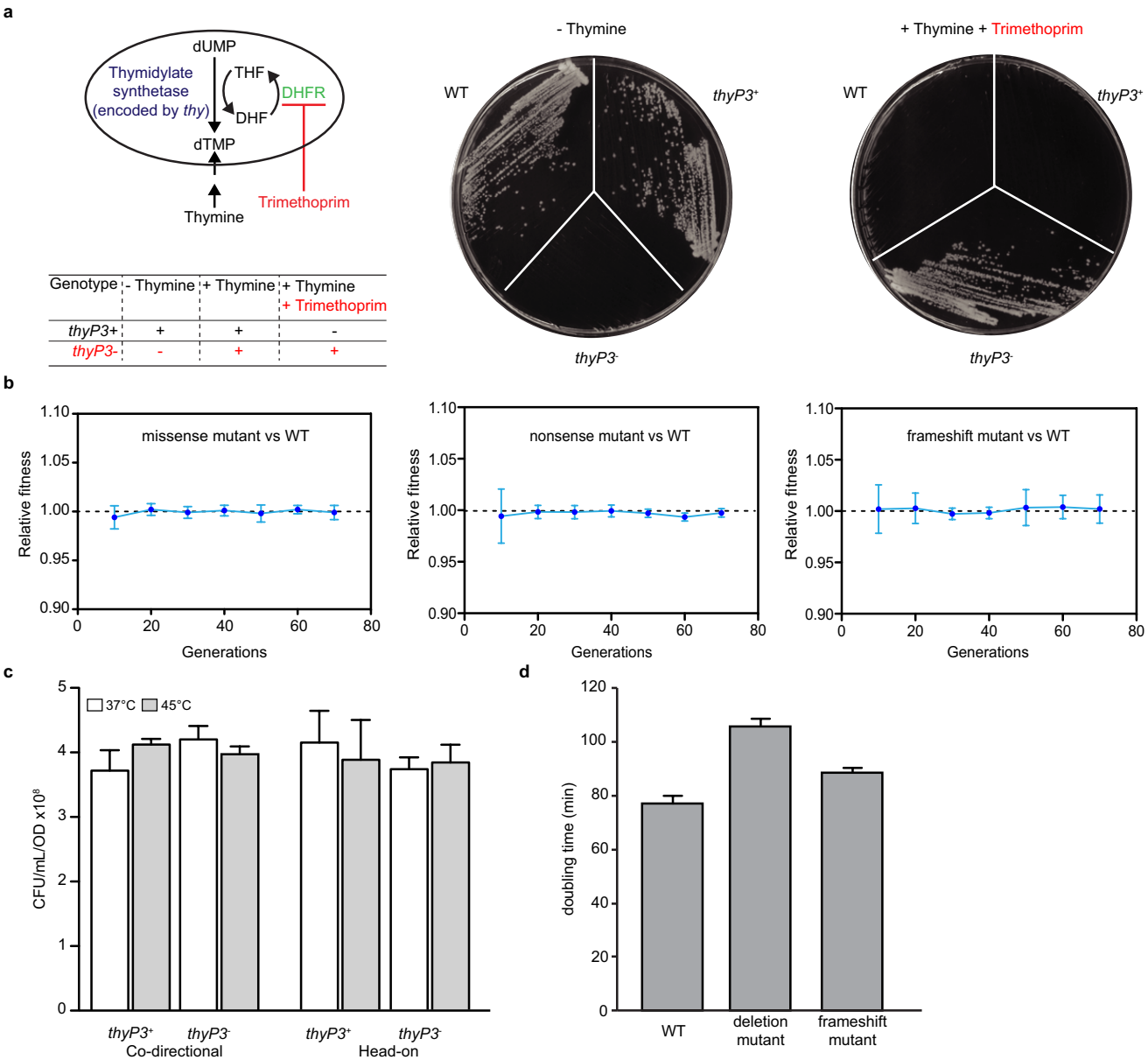
For each promoter, substitutions were estimated by pairwise comparison of the different strains using the baseml program of PAML package[44]. Baseml program uses a maximum likelihood approach to estimate nucleotide substitutions, based on an input phylogenetic tree. We used the maximum likelihood phylogenetic tree generated earlier and the substitution model was GTR. The rest of the parameters were default. Then for each promoter, mean substitutions per site were calculated and the distribution of mean pairwise substitution rates was compared between leading- and lagging-strand promoters. Mann–Whitney *U*-test was used to determine statistical significance.

For comparing the mutation rates between promoters with and without transcription factor binding, we used the population genetic parameter Watterson's estimator of theta (θ_W). Since theta (θ_W) is a population genetic parameter, it is well suited for analysing within-species sequence polymorphism and thus θ_W serves as a proxy for mutation rate of a given promoter. We calculated θ_W for the total number of mutations in the high-quality sequence alignment for each promoter across the eight strains of *B. subtilis* (Extended Data Table 2) using the DnaSP software (v.5)[45]. Promoters with experimentally validated transcription factor binding were obtained from the DBTBS database[37]. Sequence covering the +1 site, −10 and −35 elements, which includes the transcription-factor-binding site, were used for constructing the alignment as described earlier. A total of 33 different transcription factors that are experimentally validated in *B. subtilis* were used (Extended Data Table 2). Mann–Whitney *U*-test was used to determine statistical significance.

Nitrous acid mutagenesis. Nitrous acid is known to strongly deaminate purines and pyrimidines in DNA. Adenine is deaminated to hypoxanthine[46] that produces A:T to G:C transition. We subjected the wild-type and *yxlJ* (encoding hypoxanthine-DNA glycosylase)[47] mutant strains carrying the head-on *thyP3* reporter under induced transcription to nitrous acid treatment following the protocol reported previously[48]. Briefly, cells were grown in 5 ml of S7 minimal medium with 20 μg ml^{-1} thymine, 40 μg ml^{-1} tryptophan and with 1 mM IPTG for 12 h to saturation. To the saturated cultures 1 ml of 8.7 M $NaNO_2$ (nitrous acid dissolved in sodium acetate buffer pH 4.6) (Sigma-Aldrich) was added and incubated at room temperature for 60 min. As a control, cells were treated with the sodium acetate buffer in parallel. Cells were then spun down, washed and re-suspended in the growth medium and 1 ml of culture was used for determining the c.f.u. and the rest of the culture was plated on minimal plates supplemented with 20 μg ml^{-1} thymine, 40 μg ml^{-1} tryptophan, 1 mM IPTG and 5 μg ml^{-1} trimethoprim for selecting trimethoprim-resistance mutants. The same was performed for buffer-treated cells except that 0.1 ml of culture was used to determine c.f.u. and 0.1 ml was plated for selecting trimethoprim-resistant colonies. After 2 days of incubation, trimethoprim-resistance colonies appeared, and as described before the *thyP3* gene was PCR amplified and screened for $T_{-7}\rightarrow C_{-7}$ mutation. Mutation frequency was calculated by dividing the number of trimethoprim-resistant colonies by number of colonies on nonselective plate. Experiment was done in triplicate and error bars represent s.e.m. Statistical significance was obtained using Student's *t*-test.

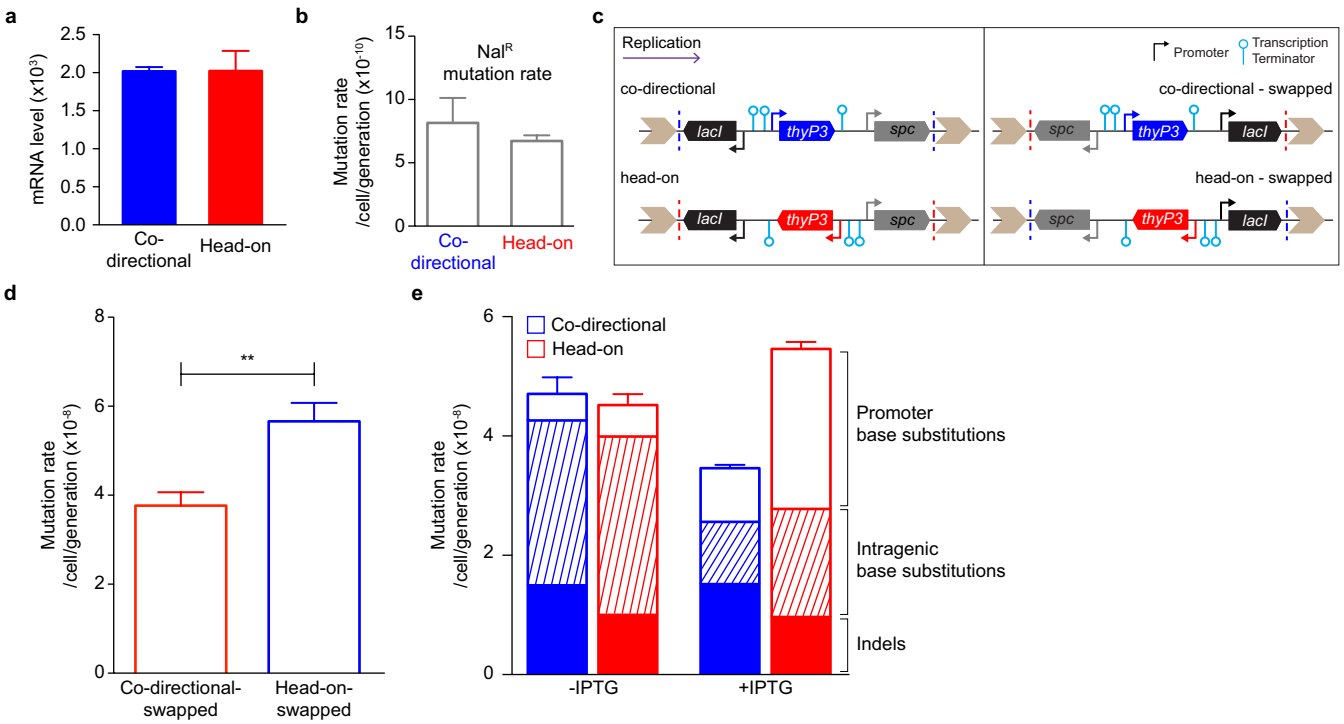
31. Vasantha, N. & Freese, E. Enzyme changes during *Bacillus subtilis* sporulation caused by deprivation of guanine nucleotides. *J. Bacteriol.* **144,** 1119–1125 (1980).
32. Janes, B. K. & Stibitz, S. Routine markerless gene replacement in *Bacillus anthracis*. *Infect. Immun.* **74,** 1949–1953 (2006).
33. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).
34. Hall, B. M., Ma, C.-X., Liang, P. & Singh, K. K. Fluctuation analysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25,** 1564–1565 (2009).
35. Lippert, M. J. *et al.* Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc. Natl Acad. Sci. USA* **108,** 698–703 (2011).
36. Ter Beek, A. *et al.* Transcriptome analysis of sorbic acid-stressed *Bacillus subtilis* reveals a nutrient limitation response and indicates plasma membrane remodeling. *J. Bacteriol.* **190,** 1751–1761 (2008).
37. Sierro, N., Makita, Y., de Hoon, M. & Nakai, K. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Res.* **36,** D93–D96 (2008).
38. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
39. Markowitz, V. M. *et al.* IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40,** D115–D122 (2012).
40. Gao, F. & Zhang, C.-T. DoriC: a database of *oriC* regions in bacterial genomes. *Bioinformatics* **23,** 1866–1867 (2007).
41. Katoh, K. & Toh, H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26,** 1899–1900 (2010).
42. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34,** W609–W612 (2006).
43. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321 (2010).
44. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24,** 1586–1591 (2007).
45. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–1452 (2009).
46. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).
47. Aamodt, R. M., Falnes, P. Ø., Johansen, R. F., Seeberg, E. & Bjørås, M. The *Bacillus subtilis* counterpart of the mammalian 3-methyladenine DNA glycosylase has hypoxanthine and 1,N^6-ethenoadenine as preferred substrates. *J. Biol. Chem.* **279,** 13601–13606 (2004).
48. Zamenhof, S. Gene unstabilization induced by heat and by nitrous acid. *J. Bacteriol.* **81,** 111–117 (1961).

© 2016 Macmillan Publishers Limited. All rights reserved



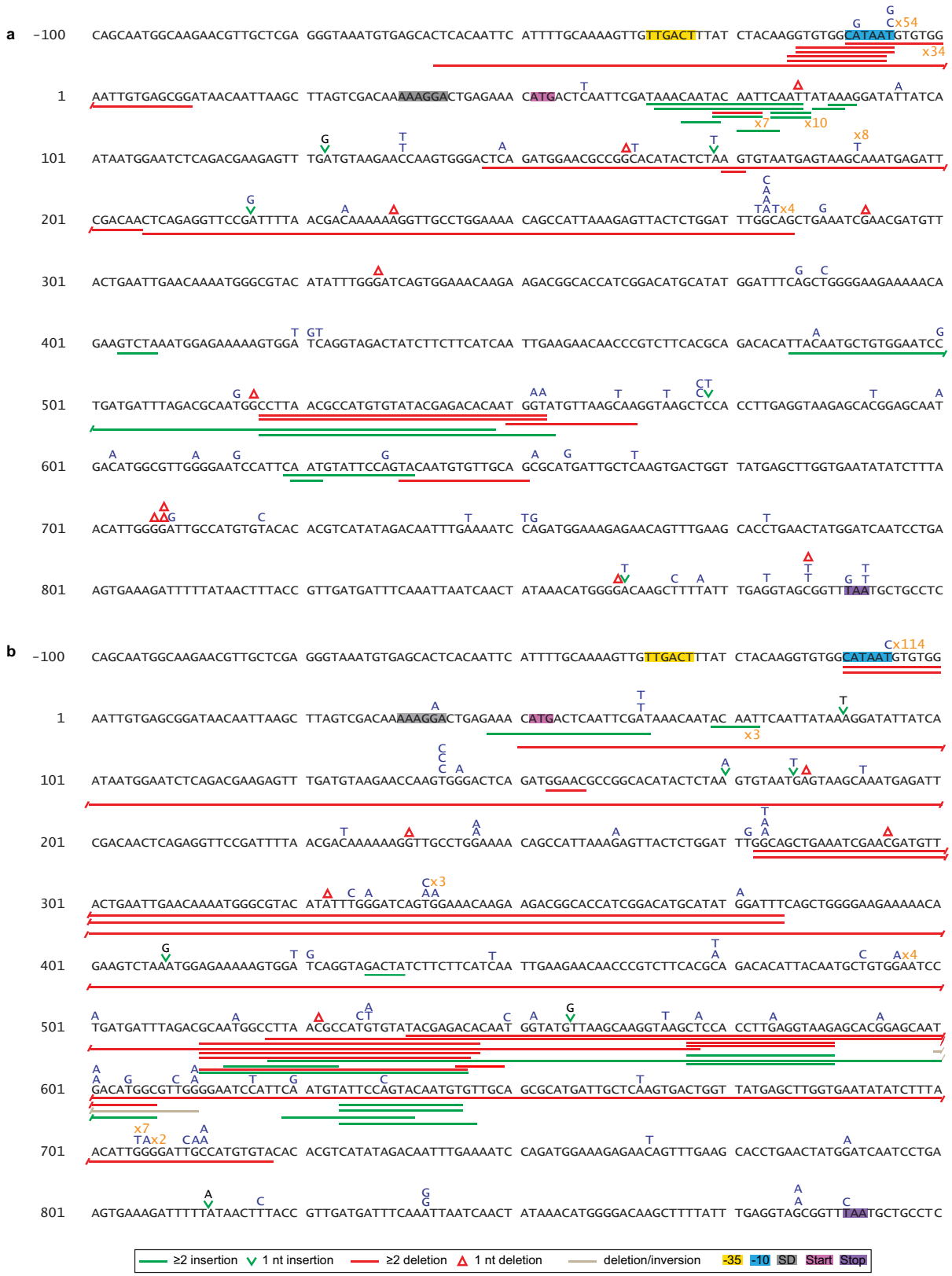
Extended Data Figure 1 | Development of a forward mutation assay that detects loss-of-function mutations in *B. subtilis*. **a**, Simplified diagram of thymidine monophosphate (dTMP) synthesis. The phage-encoded *thyP3* encodes thymidylate synthetase, which synthesizes dTMP and dihydrofolate (DHF) from dUMP and tetrahydrofolate (THF). DHF is recycled back to THF by dihydrofolate reductase (DHFR). Trimethoprim inhibits DHFR, thus blocking recycling of the essential cofactor THF and available THF is depleted by active thymidylate synthetase and cell growth is inhibited. Because cells with active thymidylate synthetase rely solely on endogenous dTMP synthesis, *thyP3*$^{+}$ cells are sensitive to trimethoprim and loss-of-function mutations in *thyP3* lead to trimethoprim resistance, which is the basis for the forward mutation assay. Viabilities of wild-type (*thyA*$^{+}$ *thyB*^{ts}), *thyP3*$^{+}$ (Δ*thyA thyB*^{ts} *thyP3*$^{+}$) and *thyP3*$^{-}$ (Δ*thyA thyB*^{ts} *thyP3*$^{-}$) cells are shown in the table and representative colonies (at 45 °C) are shown on the right. **b**, Competition between strains carrying wild-type (WT) and mutant *thyP3* in a Δ*thyA thyB*^{ts} background to determine if there is any selective pressure on different mutants during growth phase at permissive temperature (37 °C). Relative fitness (mean ± s.d.) of six replicates is shown. **c**, Shifting the temperature to 45 °C does not affect plating efficiency during selection for trimethoprim resistance. Wild-type and mutant *thyP3* cells were grown at 37 °C and plated on solid medium supplemented with IPTG plus thymine at 37 °C and 45 °C, and colony-forming units (c.f.u.) ml^{-1} optical density ($OD_{600\,nm}$)$^{-1}$ was determined. Mean ± s.d. of three replicates is shown. **d**, *thyP3* mutants have growth defects without *thyB*^{ts}. The doubling times of *thyP3* mutant (a deletion and a frameshift mutant) in the Δ*thyA* Δ*thyB* background at 37 °C are longer, indicative of growth defects in the absence of the backup gene *thyB*. Mean ± s.d. of three replicates is shown. For **b** and **d**, the mutant strains are listed in Extended Data Table 1.

© 2016 Macmillan Publishers Limited. All rights reserved



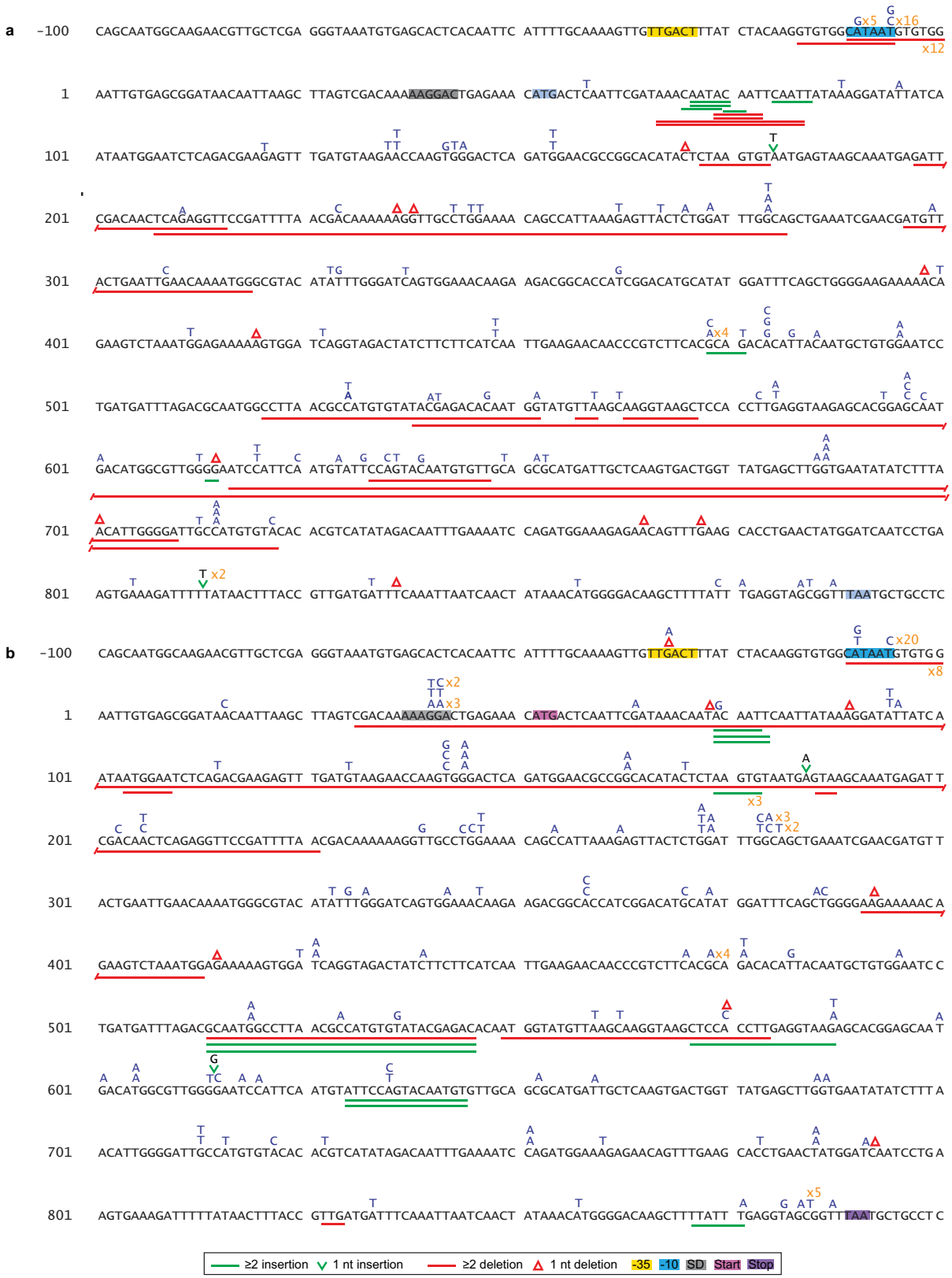
Extended Data Figure 2 | Expression level and mutation rate of *thyP3*. **a**, *thyP3* expression in co-directional and head-on orientations. Using real-time quantitative PCR, messenger RNA level of *thyP3* in the co-directional and head-on strains under induced (+IPTG) condition was measured and normalized to the reference gene *accA*. Since level of expression is similar between the strains, the observed difference in *thyP3* mutation rate between co-directional and head-on orientations (Fig. 1d) is not caused by intrinsic differences in the expression level of *thyP3*. **b**, The orientation-specific difference in *thyP3* mutation rate is not due to global increase of mutagenesis in the head-on strain. As a control to show that the increase in mutation rate is local to the *thyP3* reporter, we measured the mutation rate for resistance to nalidixic acid (Nal^R, conferred by mutations in the *gyrA* gene) in co-directional and head-on strains. Since the Nal^R mutation rates in the two strains were similar, we conclude that the observed increase in head-on mutation rate is specific to *thyP3* gene. **c**, Schematics of the co-directional and head-on *thyP3* constructs (left) and an additional control to examine the effect of the genomic context on *thyP3* mutagenesis, the neighbouring genes were swapped (right). In each construct, the *thyP3* gene is flanked by the *lacI* gene and the spectinomycin-resistance gene. The reporter constructs were integrated into the chromosome at the *amyE* locus by double crossover. The direction of replication is shown at the top. The co-directional-swapped strain was created by inverting the *lacI-thyP3-spc* from the head-on strain and the head-on-swapped construct was created by inverting the same construct from the co-directional strain. Thus the swapped constructs switch the neighbouring transcription units. The dotted lines in each construct show the swapping boundary. **d**, The mutation rate of the swapped head-on strain is still higher than the swapped co-directional strain when transcription is induced (+IPTG), indicating that the difference in mutation rate between reporter strains is not due to the direction of *thyP3* relative to its neighbouring genes. **e**, Mutation rate of co-directional and head-on *thyP3* under uninduced (−IPTG) and induced (+IPTG) transcription. The rate of each class of mutations obtained under each condition is also depicted within each bar. For **b**, **d** and **e**, mean ± s.e.m. of $n \geq 3$ independent experiments is shown. $^{**}P < 0.01$, Student's *t*-test.

© 2016 Macmillan Publishers Limited. All rights reserved



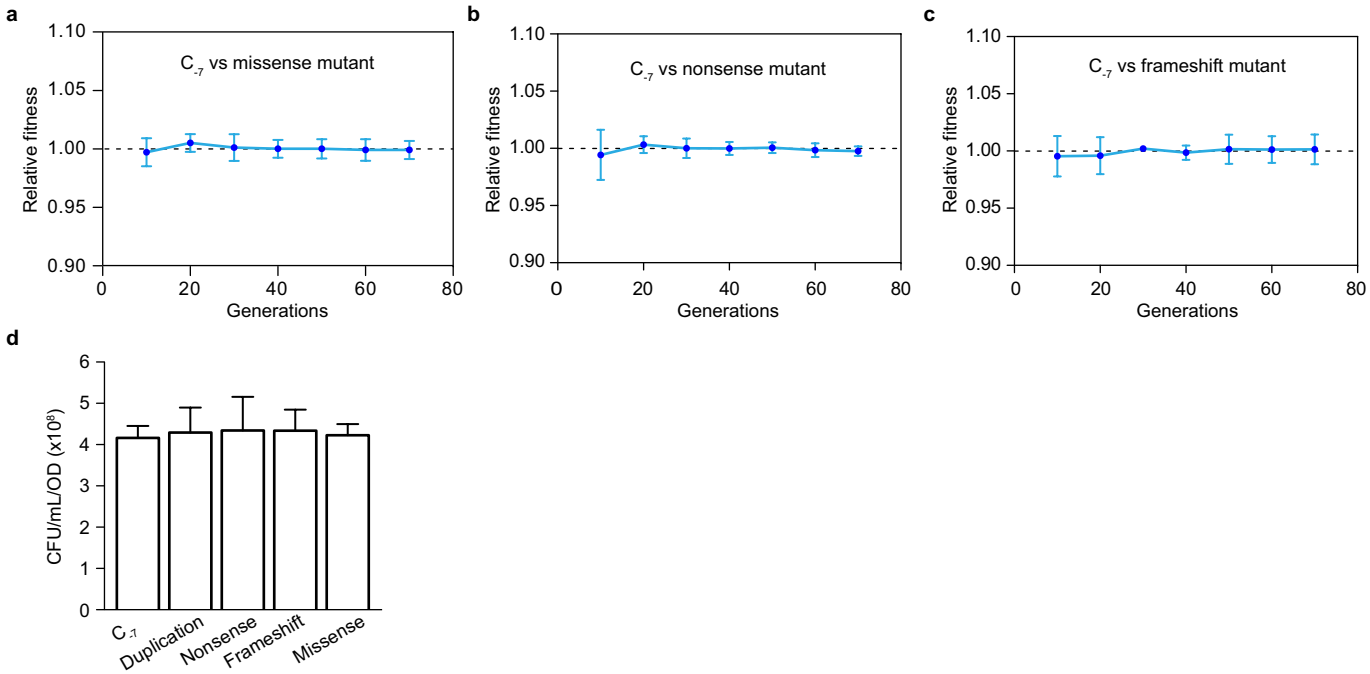
Extended Data Figure 3 | Mutation spectra of *thyP3* under induced transcription. a, **b**, Illustrations of the mutation spectra of the *thyP3* mutants obtained from fluctuation tests of co-directional ($n = 214$) (**a**) and head-on ($n = 232$) (**b**) strains when transcription is induced (+IPTG). The *thyP3* coding sequence with its promoter is shown. Sequence coordinates are indicated with reference to +1 transcription start site. The symbols used to represent different mutations are shown at the bottom, and base substitutions are shown in blue above the sequence. The numbers marked in orange next to a mutation denote the frequency. The promoter elements, Shine–Dalgarno (SD) sequence, and start and stop codons are highlighted in each spectrum.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 4 | Mutation spectra of *thyP3* under uninduced transcription. a, **b**, Illustrations of the mutation spectra of the *thyP3* mutants obtained from fluctuation tests of co-directional ($n = 163$) (**a**) and head-on ($n = 178$) (**b**) strains when transcription is not induced (−IPTG). The *thyP3* coding sequence with its promoter is shown. Sequence coordinates are indicated with reference to +1 transcription start site. The symbols used to represent different mutations are shown at the bottom, and base substitutions are shown in blue above the sequence. The numbers marked in orange next to a mutation denote the frequency. The promoter elements, Shine–Dalgarno (SD) sequence, and start and stop codons are highlighted in each spectrum.

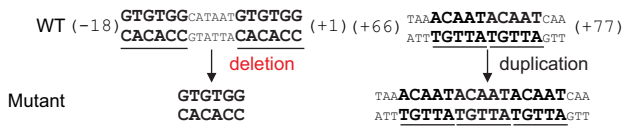
© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 5 | Absence of selection bias in *thyP3* forward mutation assay. a–c, Growth competition experiments were performed between the C_{-7} promoter mutant against the following mutants: missense mutant (**a**), nonsense mutant (**b**) and frameshift mutant (**c**). Each mutant was competed against the C_{-7} promoter mutant to check if there is a competitive disadvantage for a mutant that has a mutation within the coding sequence, which may explain the high frequency of C_{-7} mutation compared to other mutations. The results show no fitness disadvantage for any of the mutants tested, suggesting that the high frequency of C_{-7} promoter mutation is not due to a selection bias. Mean ± s.d. of six replicates is shown and mutants competed are indicated within the plot. **d**, Plating efficiency of different *thyP3* mutants. Plating efficiency was determined to check whether different classes of *thyP3* mutants have differences in their plating efficiency on trimethoprim-selection plates at 45 °C, which may explain the variation in the mutation rates and spectrum. The result shows similar plating efficiency among the different mutants, suggesting that plating efficiency does not underlie the variation in the observed mutation rates. The different mutants tested are indicated on the *x* axis. Mean ± s.d. of three replicates is shown. The mutant strains are listed in the Extended Data Table 1.

© 2016 Macmillan Publishers Limited. All rights reserved

a

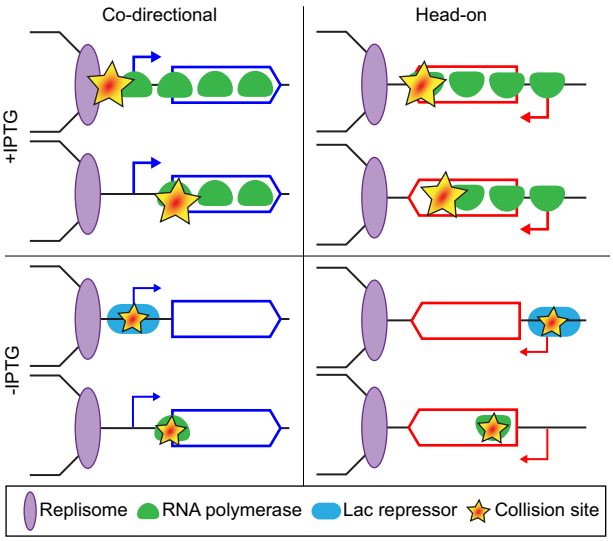


b

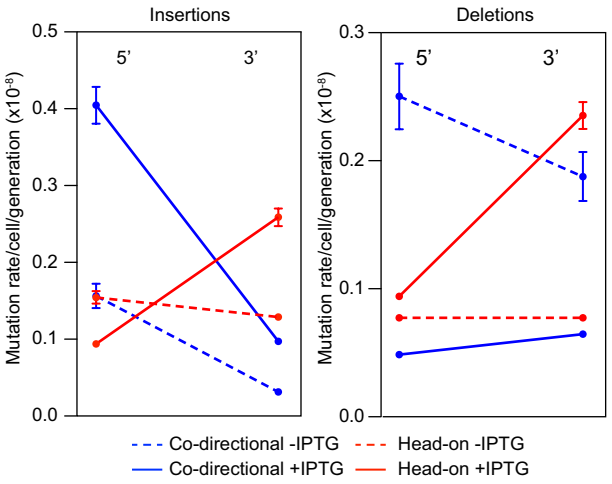
Mutation rate of indels normalized to the length of the region (x10^{-11})

Position	Co-directional -IPTG	Head-on -IPTG	Co-directional +IPTG	Head-on +IPTG
5′	0.96	0.61	1.12	0.48
3′	0.52	0.55	0.38	1.18
Promoter	10.2	5.79	15.8	1.17

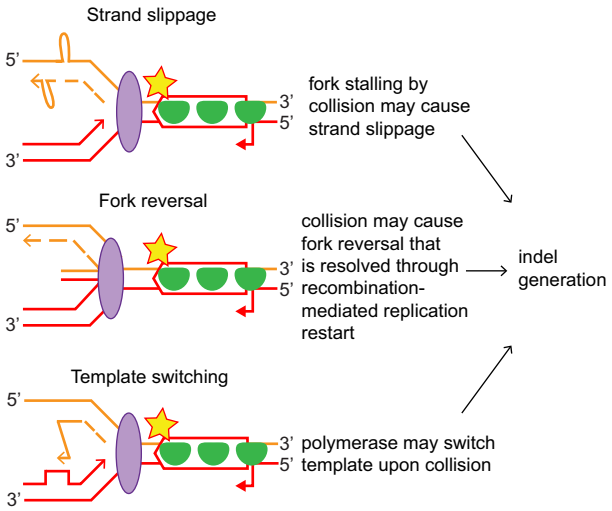
c



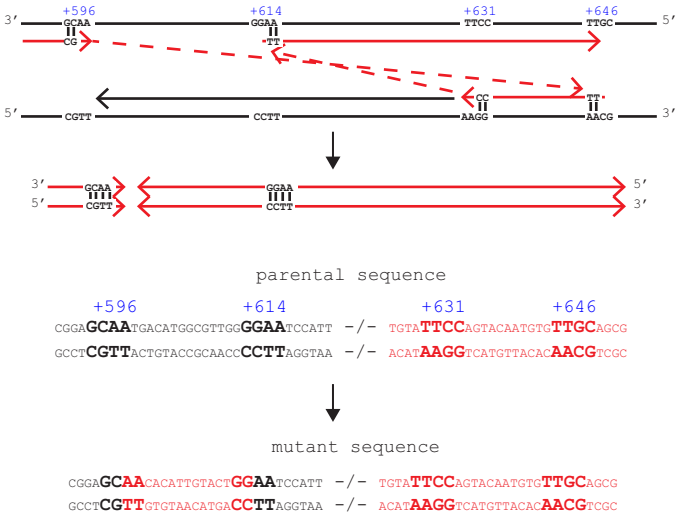
d



e

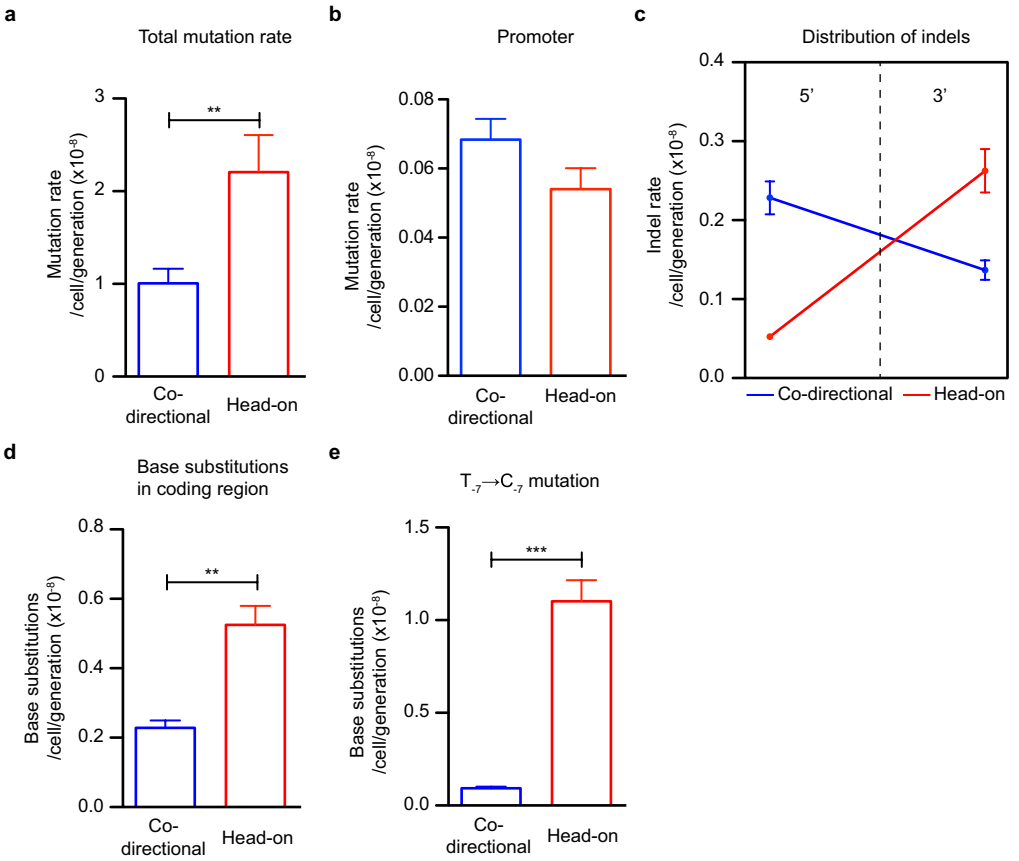


f



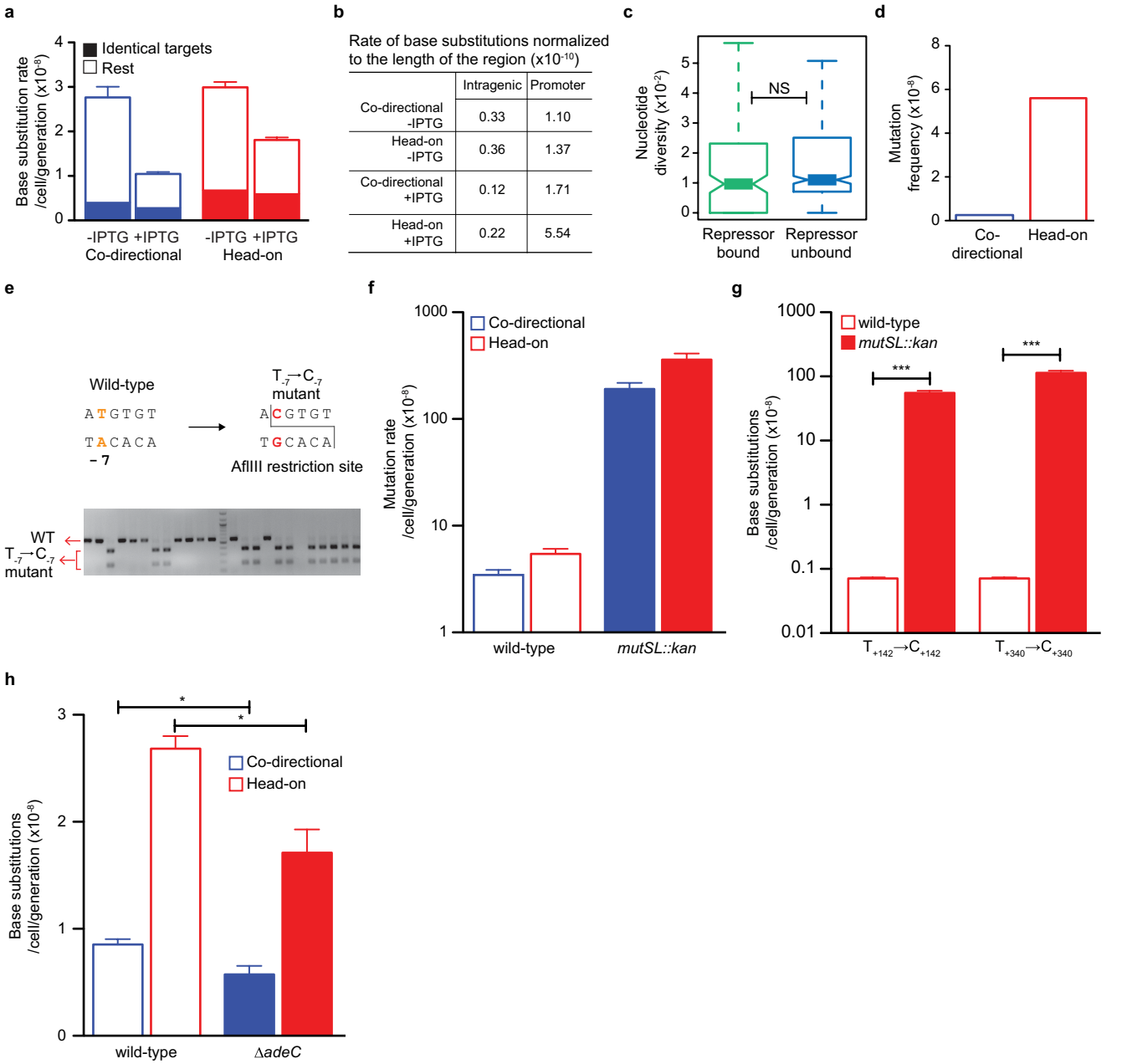
Extended Data Figure 6 | Mechanism of indel generation. a, Representative deletion and duplication events in *thyP3*. A high-frequency deletion and duplication event observed in the *thyP3* gene in co-directional and head-on strains. The sequence coordinates are denoted and repeat sequence is underlined. **b**, Table showing the mutation rate of indels (≥3 bp) in intragenic region and promoter normalized by the length of the region suggests that the localized rate of indels is higher in the promoter than the intragenic region. **c**, First encounter between replication and transcription machineries generates indels. Model describing the first-encounter hypothesis proposed on the basis of results presented in Fig. 2a–f. In co-directional orientation under induced transcription (+IPTG), when an array of RNA polymerases (RNAPs) transcribe the gene, the replisome is likely to collide with the first transcription complex at the promoter or promoter-proximal regions. By contrast, when transcription is induced in head-on orientation, the replisome encounters the first transcription complex from the 3′ end. In support of this first-encounter model, when transcription is not induced (basal level) the density of RNAP is sparse along the gene, hence the sites of collisions are altered. In addition, it is possible that under basal transcription, replisome collides with either the RNAP complex arrested at the promoter or with the Lac repressor, which may explain the relatively high frequency of deletion at the promoter. Thus the first-encounter model of replication–transcription collisions supports the idea that collisions stall replisome progression, triggering indel mutations. **d**, Mutation rate of insertions and deletions (≥3 bp) within 5′ or 3′ half of the intragenic region. Mean ± s.e.m. of $n \geq 3$ experiments is shown. **e**, Models illustrating the different pathways that can lead to generation of indels after head-on collision-induced replication stalling: slippage, fork-reversal or template switching. **f**, Illustration of a complex mutation observed in *thyP3* that is probably generated via microhomology-mediated break-induced replication. The complex mutation encompassing a deletion and insertion of an inverted region was observed under induced transcription in head-on orientation. The sequence coordinates are marked on the top with reference to the transcription start site (+1).

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 7 | Role of recombination protein RecA in collision-induced mutations. a, Mutation rates of co-directional and head-on *thyP3* strains for trimethoprim resistance in Δ*recA* background. Similar to wild type the mutation rate of head-on is higher than the co-directional strain, although the total rate of mutation is decreased in a Δ*recA* background. **b**, Rate of ≥3 bp indels at the promoter in co-directional orientation is strongly decreased in Δ*recA* cells, suggesting that indels at the promoter are mostly RecA-dependent. **c**, Intragenic distribution of ≥3 bp indels in Δ*recA* is similar to the distribution observed for wild type (Fig. 2f), thus suggesting that RecA is not necessary for the collision-induced indels within the coding region. **d**, Mutation rate of base substitutions in Δ*recA* cells is higher in head-on than co-directional orientation. **e**, The rate of $T_{-7}\rightarrow C_{-7}$ mutation is higher in head-on relative to co-directional orientation in Δ*recA* cells, thus promoter substitutions can occur at a higher rate independent of recombination-mediated repair. All the fluctuation tests in a Δ*recA* background were performed under inducing conditions (+IPTG). Mean ± s.e.m. of $n \geq 3$ experiments is shown. $^{**}P < 0.01$; $^{***}P < 0.001$; Student's *t*-test.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 8 | Base substitutions and the role of mismatch repair and enzymatic adenine deamination. a, IPTG induction does not affect the base substitution rate in the coding region of *thyP3* when considering identical target sites, indicating that collisions may not be a major source of these mutations. In yeast, it was shown that transcription-associated mutagenesis is proportional to the level of transcription[10]. In *B. subtilis*, the total rate of base substitutions in the coding region decreases upon IPTG induction, which could be due to an unidentified transcription-dependent mutation-correction mechanism, or due to increase of target size of base substitutions in the coding sequence in uninduced (basal) transcription. **b**, Table showing the rates of base substitutions in the coding region and promoter of *thyP3* normalized by length of the region. Localized substitution rates are higher in the promoter than coding sequence, thus suggesting that collision has a more drastic effect on promoter substitutions. **c**, Comparative genomic analysis of mutation rates of promoters with and without repressor binding. Nucleotide diversity per site (theta) was calculated for each promoter across different strains of *B. subtilis*. The comparison shows no significant difference in nucleotide diversity between repressor-bound promoters and the rest of the promoters, indicating that repressor binding may not affect the substitution rate of a promoter. Whole genomes and the repressors analysed are listed in Extended Data Table 2. NS, not significant, $P > 0.05$; Mann–Whitney U-test. **d**, The mutation frequency of $T_{-7}\rightarrow C_{-7}$ mutation is higher in head-on than co-directional orientation in *E. coli*. The mutation frequency was calculated here from the plasmid-based forward mutation assay data reported previously[26]. **e**, The restriction-digestion-based assay to screen for $T_{-7}\rightarrow C_{-7}$ mutation. Wild-type promoter sequence does not have an AflIII restriction site, whereas the promoter $T_{-7}\rightarrow C_{-7}$ mutation will be digested by AflIII, which is illustrated by a representative agarose gel. **f**, Mismatch repair mutant (*mutSL::kan*) shows an expected increase (~60-fold) in total mutation rate of *thyP3* in both co-directional and head-on orientation compared to wild-type. The mutation rates of the wild-type strains are presented in Fig. 1d. **g**, Mismatch repair mutant shows a marked ~1,000-fold increase in mutation rate of T→C substitution hotspots within the coding sequence of head-on *thyP3*, indicating that mismatch repair corrects T→C substitution within the coding sequence. **h**, Deletion of *adeC* gene encoding adenine deaminase modestly reduces the mutation rate of $T_{-7}\rightarrow C_{-7}$ substitution in both co-directional and head-on orientation compared with wild type. For **f**–**h**, mean ± s.e.m. of $n \geq 3$ experiments is shown. $^{*}P < 0.05$; $^{***}P < 0.001$; Student's *t*-test.

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 1 | Strains and plasmids

a

Name	Genotype	Source
JDW437	(wild-type 168) *trpC2*	Lab stock
JDW941	151 *phi3T*	Ronald Yasbin
JDW942	168 *thyA⁻ thyB⁻*	Ronald Yasbin
JDW1297	PY79 *mutSL::kan*	Lyle Simmons
JDW1543	168 Δ*thyA*	This work
JDW1544	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) *spc*	This work
JDW1563	168 Δ*thyA amyE::P$_{spank}$-thyP3* (co-directional) *spc*	This work
JDW1711	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) *spc, lacA::P$_{pen}$-spoVG-lacZ*	This work
JDW1814	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) *spc, mutSL::kan*	This work
JDW1900	168 Δ*thyA amyE::spc-P$_{spank}$-thyP3* (head-on) *lacI*	This work
JDW1901	168 Δ*thyA amyE::spc-P$_{spank}$-thyP3* (co-directional) *lacI*	This work
JDW2054	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $G_{+473}\rightarrow A_{+473}$ *spc*	This work
JDW2057	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $T_{-7}\rightarrow C_{-7}$ *spc*	This work
JDW2185	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $T_{+331}\rightarrow C_{+331}$ *spc*	This work
JDW2190	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $G_{+473}\rightarrow A_{+473}$ *spc, lacA::P$_{pen}$-spoVG-lacZ*	This work
JDW2192	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $T_{-7}\rightarrow C_{-7}$ *spc, lacA::P$_{pen}$-spoVG-lacZ*	This work
JDW2266	168 Δ*yqjH*	BGSC
JDW2284	168 Δ*yxlJ*	BGSC
JDW2288	168 Δ*recA*	BGSC
JDW2491	168 *thyA⁻ thyB⁻ amyE::P$_{spac(hy)}$-thyP3* $\Delta_{+102\text{-}145}$ deletion *spc*	This work
JDW2492	168 *thyA⁻ thyB⁻ amyE::P$_{spac(hy)}$-thyP3* TT_{+124} insertion *spc*	This work
JDW2501	168 Δ*adeC*	BGSC
JDW2529	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) *spc,* Δ*yqjH*	This work
JDW2530	168 Δ*thyA amyE::P$_{spank}$-thyP3* (co-directional) *spc,* Δ*yqjH*	This work
JDW2547	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) *spc,* Δ*adeC*	This work
JDW2548	168 Δ*thyA amyE::P$_{spank}$-thyP3* (co-directional) *spc,* Δ*adeC*	This work
JDW2598	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) *spc,* Δ*recA*	This work
JDW2612	168 Δ*thyA amyE::P$_{spank}$-thyP3* (co-directional) *spc,* Δ*recA*	This work
JDW2697	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) *spc,* Δ*yxlJ*	This work
JDW2746	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $G_{+332}\rightarrow A_{+332}$ *spc*	This work
JDW2747	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $G_{+332}\rightarrow A_{+332}$ *spc, lacA::P$_{pen}$-spoVG-lacZ*	This work
JDW2748	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $+1G_{+335}$ *spc*	This work
JDW2749	168 Δ*thyA amyE::P$_{spank}$-thyP3* (head-on) $+1G_{+335}$ *spc, lacA::P$_{pen}$-spoVG-lacZ*	This work

b

Name	Genotype	Source
pDR90	*amyE::P$_{spac(hy)}$ amp spc*	David Rudner
pDR110	*amyE::P$_{spank}$ amp spc*	David Rudner
pJW299	pEX44/I-*Sce*I site *amp cat*	Lab stock
pJW331	pDR90/*amyE::P$_{spac(hy)}$-thyP3* (head-on) *amp spc*	This work
pJW395	pJW299/Δ*thyA* I-*Sce*I site *amp cat*	This work
pJW396	pDR110/*amyE::P$_{spank}$-thyP3* (head-on) *amp spc*	This work
pJW397	pDR110/*amyE::P$_{spank}$-thyP3* (co-directional) *amp spc*	This work
pJW417	pEX44/*lacA::P$_{pen}$-spoVG-lacZ amp cat*	This work
pJW430	pDR110/*amyE::spc-P$_{spank}$-thyP3* (head-on) *lacI amp*	This work
pJW431	pDR110/*amyE::spc-P$_{spank}$-thyP3* (co-directional) *lacI amp*	This work

a, Bacterial strains used in this study. **b**, Plasmids used in this study. BGSC, Bacillus Genetic Stock Center (http://www.bgsc.org).

© 2016 Macmillan Publishers Limited. All rights reserved

Extended Data Table 2 | Primers, whole-genome sequences and transcriptional regulators

a

Name	Sequence 5'→3'
oJW760	GGTGTCGACATGACTCAATTCGATAAACAA
oJW761	AATGGCATGCCAATATTTCACCAATTTCAT
oJW785	GTATGAATTCCAATATTTCACCAATTTCAT
oJW1011	GCGGATAACAATTTCACACAGGGTCTTCTTGTTTCCACTGAT
oJW1013	GCGGATAACAATTTCACACAGG CAATATTTCACCAATTTCAT
oJW1052	GGTAGAATTCACGTTATGGTTAAGATTCAA
oJW1053	AATGCTCGAGTATCCTTCTTTCATTTTCAG
oJW1054	GGTACTCGAGTAGCAGGTATCCTAATTTCA
oJW1055	AATGGGATCCCAGTCCAAATGACAATCTAT
oJW1137	ATTGGCATGCTCGACTCTCTAGCTTGAG
oJW1199	TGGTGTCAAAAATAACTCGACCTTCGATATGGGCGGATTCTT
oJW1200	GAATCCGCCCATATCGAAGGTCGAGTTATTTTTGACACCA
oJW1201	TGATGTTTGAGTCGGCTGATAGGGAAAAGGTGGTGAACTAC
oJW1202	GTAGTTCACCACCTTTTCCCTATCAGCCGACTCAAACATCAAA
oJW1203	GGCTAAGAGAACAAGGAGGAGACGGTGGAAACGAGGTCATCATTT
oJW1204	ATGACCTCGTTTCCACCGTCTCCTCCTTGTTCTCTTAGCC
oJW1213	CATAAAGGCTAGGGATAACAGGGTAATCCGCTCACAATTCCACACAAC
oJW1214	GCAGACGTTGCCATATCCAATTCAAGCTGGGGATCCTAGAAGCT
oJW1215	CTTCTAGGATCCCCAGCTTGAATTGGATATGGCAACGTCTGCCC
oJW1217	CAGAGGTTCCGATTTTAAC
oJW1218	TCAATTCAGTAACATCGTTC
oJW1221	GCTTCAGGATGATATTTACAA
oJW1222	CAGGTGTTCGATATAATCAAG
oJW1335	GTAAAACGACGGCCAGTGCGTTTCGGTGATGAAGAT
oJW1336	ATTAAAAACTGGTCTGATCGCTATGCAAGGGTTTATTGTT
oJW1337	AACAATAAACCCTTGCATAGCGATCAGACCAGTTTTTAAT
oJW1338	AGGAAATCCATTATGTACTATTTAGTACGCCTCTTTTCTTTTC
oJW1339	GAAAAGAAAAGAGGCGTACTAAATAGTACATAATGGATTTCCT
oJW1902	CCTGACTGGGAAGAGGATGACG
oJW1903	TCAGCTTTCATGGCTATCATTGAAC
oJW1904	CTGGCTGGAAATACGCTTCTCG
oJW1905	GATCAACGACGCTCAAGAGCTCA
oJW1906	GGACTGTCCGCGTCGTTACGT
oJW1907	GCTTCCTCGCTCCCTTGGG
oJW2008	GGCATGAGCCTGGGCATGTG
oJW2009	CTCCGTCTGCGTTTCGCAGTTC

b

Bacillus genomes	NCBI_accession
Bacillus subtilis subtilis 168	NC_000964.3
Bacillus subtilis subtilis BSP1	CP003695
Bacillus subtilis QB928	CP003783.1
Bacillus subtilis 6051HGW	CP003329
Bacillus subtilis spizizenii W23	NC_014479.1
Bacillus subtilis subtilis RO-NN-1	CP002906
Bacillus subtilis spizizenii TU-B-10	NC_016047
Bacillus subtilis BSn5	NC_014976.1
Bacillus amyloliquefaciens FZB42	NC_009725.1

c

Regulator name	Function
AbrB	transcriptional regulator for transition state genes
AhrC	arginine repressor
AraR	transcriptional repressor of the ara regulon (LacI family)
BkdR	transcriptional regulator
CcpA	transcriptional regulator (LacI family)
CodY	transcriptional repressor CodY
ComA	two-component response regulator
ComK	competence transcription factor (CTF)
CtsR	transcriptional regulator
DegU	two-component response regulator
Fnr	transcriptional regulator (FNR/CAP family)
Fur	transcriptional regulator for iron transport and metabolism
GlnR	transcriptional regulator (nitrogen metabolism)
GltC	transcriptional regulator (LysR family)
GltR	transcriptional regulator (LysR family)
Hpr	transcriptional regulator Hpr
HrcA	heat-inducible transcription repressor
IolR	transcriptional regulator (DeoR family)
LevR	transcriptional regulator (NifA/NtrC family)
LexA	transcriptional repressor of the SOS regulon
MntR	manganese transport transcriptional regulator
Mta	transcriptional regulator (MerR family)
PerR	transcriptional regulator (Fur family)
PucR	transcriptional regulator of the purine degradation operon
PurR	pur operon repressor
ResD	two-component response regulator
RocR	transcriptional regulator (NtrC/NifA family)
SinR	transcriptional regulator for post-exponenetial-phase-response
Spo0A	master regulator of sporulation
SpoIIID	transcriptional regulator of mother cell gene expression
TnrA	nitrogen sensing transcriptional regulator
Xre	Phage PBSX transcriptional regulator
Zur	transcriptional regulator (Fur family)

a, Primers used in this study. **b**, Whole-genome sequences used in this study. **c**, Transcriptional regulators analysed in this study.

© 2016 Macmillan Publishers Limited. All rights reserved

LETTER

doi:10.1038/nature18324

Allosteric coupling from G protein to the agonist-binding pocket in GPCRs

Brian T. DeVree[1]*, Jacob P. Mahoney[1]*, Gisselle A. Vélez-Ruiz[1], Soren G. F. Rasmussen[2], Adam J. Kuszak[1], Elin Edwald[1], Juan-Jose Fung[2], Aashish Manglik[2], Matthieu Masureel[2], Yang Du[2], Rachel A. Matt[2], Els Pardon[3], Jan Steyaert[4], Brian K. Kobilka[2] & Roger K. Sunahara[1,5]

G-protein-coupled receptors (GPCRs) remain the primary conduit by which cells detect environmental stimuli and communicate with each other[1]. Upon activation by extracellular agonists, these seven-transmembrane-domain-containing receptors interact with heterotrimeric G proteins to regulate downstream second messenger and/or protein kinase cascades[1]. Crystallographic evidence from a prototypic GPCR, the β_2-adrenergic receptor (β_2AR), in complex with its cognate G protein, Gs, has provided a model for how agonist binding promotes conformational changes that propagate through the GPCR and into the nucleotide-binding pocket of the G protein α-subunit to catalyse GDP release, the key step required for GTP binding and activation of G proteins[2]. The structure also offers hints about how G-protein binding may, in turn, allosterically influence ligand binding. Here we provide functional evidence that G-protein coupling to the β_2AR stabilizes a 'closed' receptor conformation characterized by restricted access to and egress from the hormone-binding site. Surprisingly, the effects of G protein on the hormone-binding site can be observed in the absence of a bound agonist, where G-protein coupling driven by basal receptor activity impedes the association of agonists, partial agonists, antagonists and inverse agonists. The ability of bound ligands to dissociate from the receptor is also hindered, providing a structural explanation for the G-protein-mediated enhancement of agonist affinity, which has been observed for many GPCR–G-protein pairs. Our data also indicate that, in contrast to agonist binding alone, coupling of a G protein in the absence of an agonist stabilizes large structural changes in a GPCR. The effects of nucleotide-free G protein on ligand-binding kinetics are shared by other members of the superfamily of GPCRs, suggesting that a common mechanism may underlie G-protein-mediated enhancement of agonist affinity.

Sequencing of the human genome revealed the magnitude of the GPCR superfamily, identifying over 800 genes encoding GPCRs, making this class of receptors the third-largest gene family[3]. Despite the varying nature of the chemical stimuli, which range from photons to small-molecule odorants and hormones to larger peptides and proteins, the generation of G-protein-mediated signals proceeds by a common mechanism. After activation, the receptor engages a heterotrimeric G protein and catalyses release of GDP from the G protein α-subunit (Gα). Intracellular GTP then binds the nucleotide-free G protein, allowing it to regulate downstream effectors (adenylyl cyclase, phospholipase C, ion channels, and so on) to elicit cellular responses[4]. We recently used X-ray crystallography[2], hydrogen–deuterium exchange mass spectrometry[5] and electron microscopy[6] to characterize an agonist–GPCR–G-protein ternary complex in the absence of nucleotide. These studies revealed dramatic conformational changes in the G protein that are stabilized by binding to agonist-activated receptor and provided insight into the mechanism by which GPCRs bind G proteins to promote nucleotide exchange. Here, we suggest an explanation for the allosteric communication that links the nucleotide-binding site on the G protein to the hormone-binding site on the receptor, with a focus on conformational changes in the extracellular face of the receptor that alter access to the hormone-binding site.

GPCR–G-protein interactions have historically been monitored using radioligand binding assays. Observations as early as the 1970s

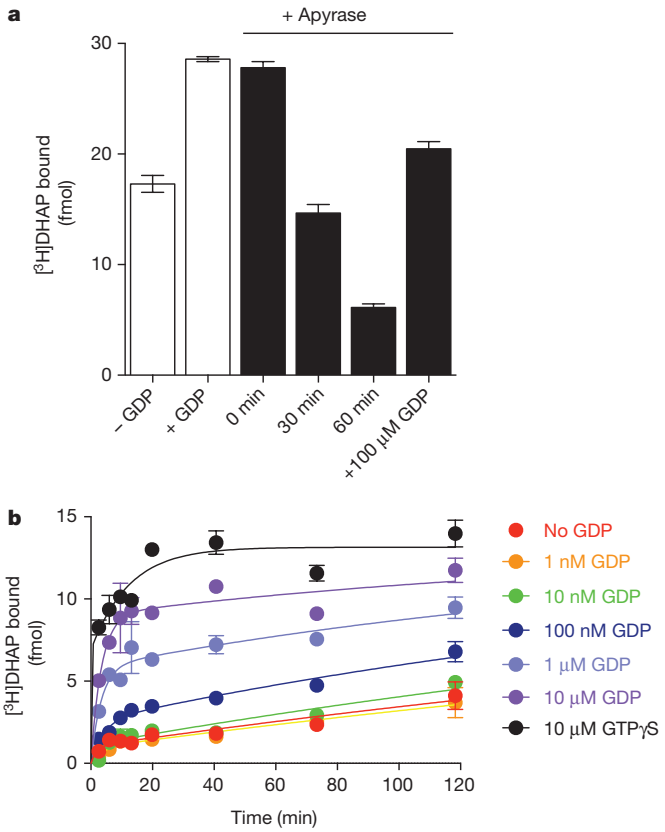


Figure 1 | Guanine nucleotides influence antagonist binding to β_2AR•Gs complexes. a, Binding of 2 nM [^{3}H]DHAP to β_2AR•Gs in the absence or presence of GDP. Addition of apyrase to GDP-bound β_2AR•Gs led to a progressive decrease in [^{3}H]DHAP binding over time, which could be restored with excess GDP. **b**, Addition of increasing concentrations of GDP enhances the rate and extent of [^{3}H]DHAP binding to apyrase-treated β_2AR•Gs complexes. **a**, Data are shown as mean ± standard error of the mean (s.e.m.) from $n = 3$ independent experiments performed in duplicate. **b**, Data are representative of three independent experiments.

[1]Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [2]Department of Cellular and Molecular Physiology, Stanford University, Palo Alto, California 94305, USA. [3]Structural Biology Research Center, VIB, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium. [4]Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium. [5]Department of Pharmacology, University of California San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA.
*These authors contributed equally to this work.

© 2016 Macmillan Publishers Limited. All rights reserved

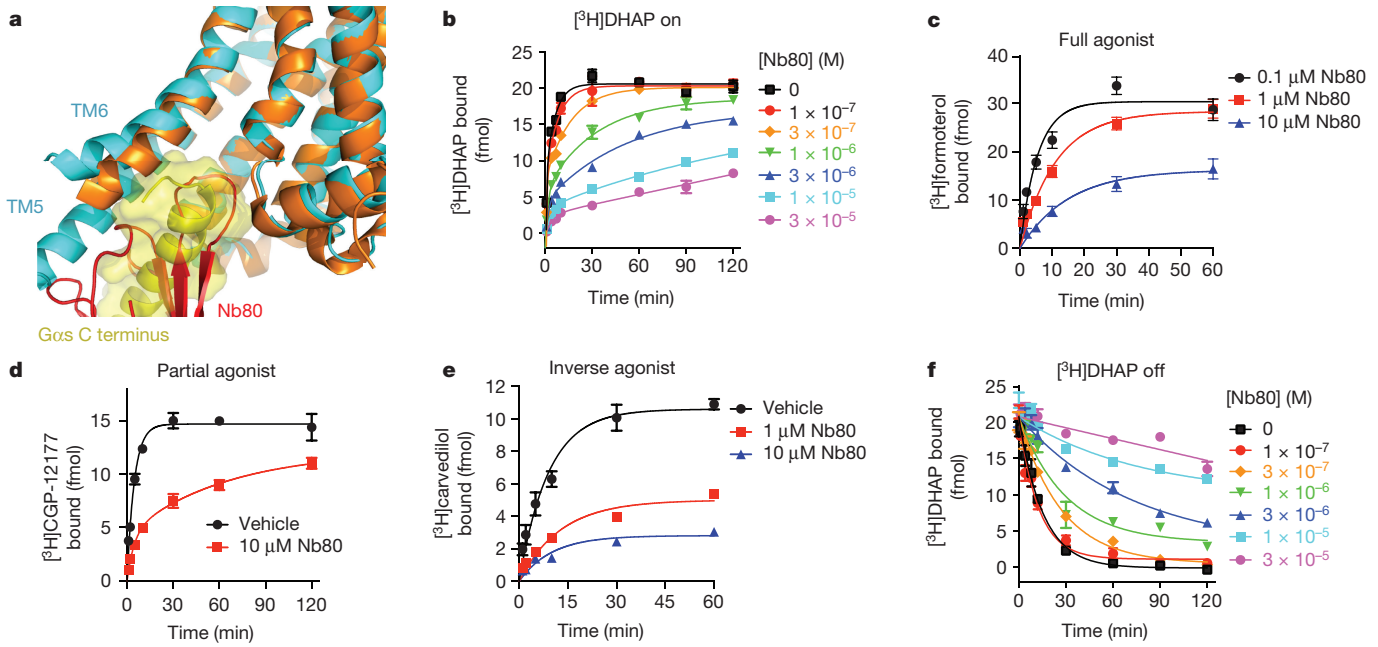


Figure 2 | Trapping active-state β_2AR with Nb80 slows both antagonist and agonist association. a, Nb80 (red) mimics G protein (yellow) in both its binding site and the β_2AR conformation it stabilizes. The structure of Nb80-bound β_2AR (Protein Data Bank (PDB) accession 3P0G) is shown in orange, Gs-bound β_2AR (PDB accession 3SN6) in cyan. **b**, Pre-incubation of the β_2AR with increasing concentrations of Nb80 progressively slows association of neutral antagonist [^{3}H]DHAP to the β_2AR. **c**–**e**, Nb80 also slows association of full agonist [^{3}H]formoterol (**c**), partial agonist [^{3}H]CGP12177 (**d**), and inverse agonist [^{3}H]carvedilol (**e**) to the β_2AR. **f**, Nb80 stabilizes the closed, active conformation and slows [^{3}H]DHAP dissociation from the β_2AR in a concentration-dependent manner. **b**, **f**, Data are representative of three independent experiments. All other data are specific binding, shown as mean ± s.e.m. from $n = 3$ independent experiments performed in duplicate.

suggested that G-protein coupling enhances agonist affinity for the receptor, and can be abolished by uncoupling the G protein from the receptor with guanine nucleotides[7]. These and other data formed the basis for the ternary complex model of agonist–receptor–G-protein interactions[8,9]. In this paradigm, the active state of the receptor is stabilized by both the agonist and G protein, and enhancement of agonist affinity arises owing to the positive cooperativity between agonist and G protein. However, using purified β_2AR•Gs complexes, we observed peculiar binding characteristics of the antagonist [^{3}H]dihydroalprenolol ([^{3}H]DHAP) to β_2AR (Fig. 1a). As illustrated, addition of GDP increases the observed binding of a saturating concentration of [^{3}H]DHAP, whereas removal of GDP using a nucleotide lyase, apyrase, decreases [^{3}H]DHAP binding. The apyrase-mediated decrease in [^{3}H]DHAP binding is reversed upon addition of excess GDP, suggesting that the decrease is indeed due to the formation of nucleotide-free β_2AR•Gs complexes. Removal of GDP from the β_2AR•Gs complex relies on the constitutive activity of β_2AR and the rapid hydrolysis (by apyrase) of GDP released from the α-subunit of Gs, Gsα. The nucleotide-free status of Gsα in these β_2AR•Gs complexes was confirmed by rapid [^{35}S]GTPγS binding kinetics (Extended Data Fig. 1)[10]. The observed deficit in [^{3}H]DHAP binding to nucleotide-free β_2AR•Gs is the result of slower [^{3}H]DHAP association (Fig. 1b and Extended Data Fig. 2). GDP enhances [^{3}H]DHAP association in a concentration-dependent manner, with similar effects achieved by complete β_2AR•Gs uncoupling with GTPγS. Although nucleotides do not significantly affect the affinity (dissociation constant (K_d)) of [^{3}H]DHAP, their modulatory capacity is γ-phosphate-dependent since GTPγS is approximately tenfold more potent than GDP (Extended Data Fig. 3). Thus, β_2AR bound to nucleotide-free G protein adopts a conformation characterized by restricted access to the hormone-binding site.

Crystallographic and pharmacological evidence suggests that the active conformation of the β_2AR is stabilized by nucleotide-free Gs or by a single-chain camelid antibody raised against agonist-bound β_2AR (nanobody Nb80) (Fig. 2a)[2,11,12]. As illustrated in Fig. 2b (and Extended Data Fig. 4a), Nb80 stabilizes a conformation of the β_2AR that restricts [^{3}H]DHAP association, similar to nucleotide-free Gs. Importantly, Nb80 also slows the association of full agonist, [^{3}H]formoterol (Fig. 2c), as well as partial agonist, [^{3}H]CGP-12177 (Fig. 2d). These data suggest that in the nucleotide-free Gs- or Nb80-stabilized active state, the β_2AR adopts a closed conformation impairing access to the orthosteric ligand-binding site, regardless of the cooperativity of the orthosteric ligand with the G protein. These data are in line with our previous observation that the β-adrenergic receptor inverse agonist ICI-118,551 blocks the formation of β_2AR•Gs complexes, but is unable to disrupt preformed complexes[10]. Nb80 also impairs binding of inverse agonist [^{3}H]carvedilol to the β_2AR by modestly decreasing the observed association rate (Fig. 2e) but dramatically decreasing total binding, suggesting that Nb80 and [^{3}H]carvedilol do not simultaneously occupy β_2AR.

Agonist-promoted G-protein engagement and subsequent nucleotide loss would be expected to stabilize the active, closed receptor conformation, thus trapping the agonist in the orthosteric site and enhancing its observed affinity. Indeed, uncoupling G protein from receptor using the GTP analogue GppNHp has been shown to accelerate agonist dissociation from the β_2AR[13]. Such agonist–G-protein cooperativity is not predicted for neutral antagonists like alprenolol, which do not stimulate G-protein coupling and thus should not stabilize the closed conformation. However, we have previously demonstrated that Gs can be 'forced' to form a complex with the β_2AR bound to the antagonist alprenolol[10], provided that free nucleotide is removed, indicating that antagonist-bound β_2AR retains enough basal activity to engage Gs. Consistent with this model, Fig. 2f and Extended Data Fig. 4b clearly illustrate a progressive slowing of [^{3}H]DHAP (or [^{3}H]CGP-12177, data not shown) dissociation in response to increasing Nb80 concentrations, suggesting that Nb80-mediated stabilization of the closed, active receptor conformation can trap [^{3}H]DHAP in the orthosteric-binding site.

Analysis of access to the hormone-binding sites in inactive- and active-state β_2AR structures provides a structural rationale for the slowing of agonist and antagonist association (Fig. 3, Extended Data Fig. 6

© 2016 Macmillan Publishers Limited. All rights reserved

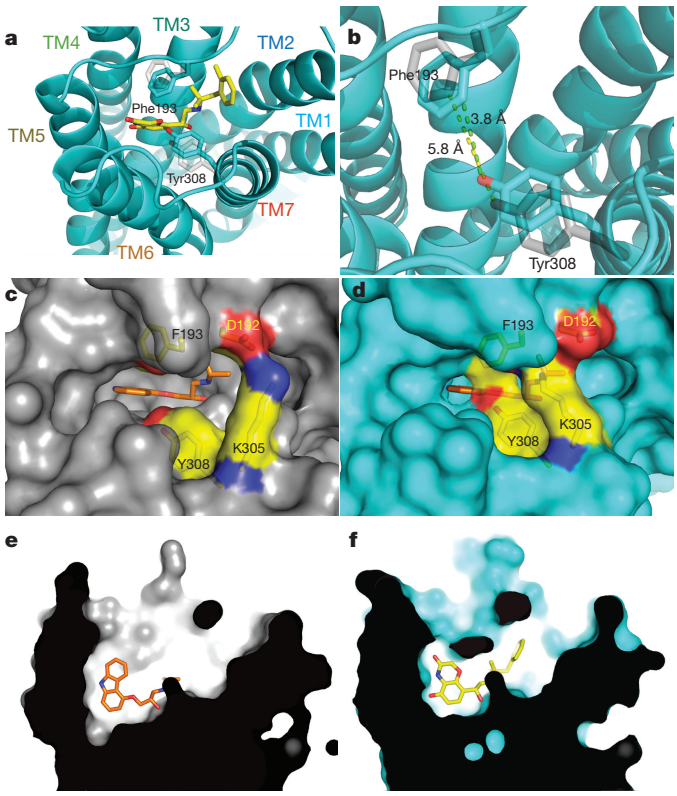


Figure 3 | Activation of the β_2AR closes the hormone-binding site. **a**, Stabilization of the β_2AR active conformation by Gs (or Nb80) brings the side chains of Phe193^{ECL2} and Tyr308$^{7.35}$ closer to one another compared to their positions in structures in the absence of G protein. **b**, Closer view of the orthosteric site, highlighting Phe193^{ECL2} and Tyr308$^{7.35}$. Distances (in Å) between the hydroxyl on Tyr308$^{7.35}$ and 2-carbon on the phenyl ring of Phe193^{ECL2} are indicated. **c**, **d**, A surface view comparing the extracellular face of β_2AR in inactive (**c**) or active (**d**) conformations, showing how G-protein-stabilized structural rearrangements occlude the hormone-binding site in the active state. **e**, **f**, Cutaway view illustrating closure of the hormone-binding site around the bound agonist in the active state. The inverse agonist carazolol is shown in orange, the agonist BI-167107 is shown in yellow.

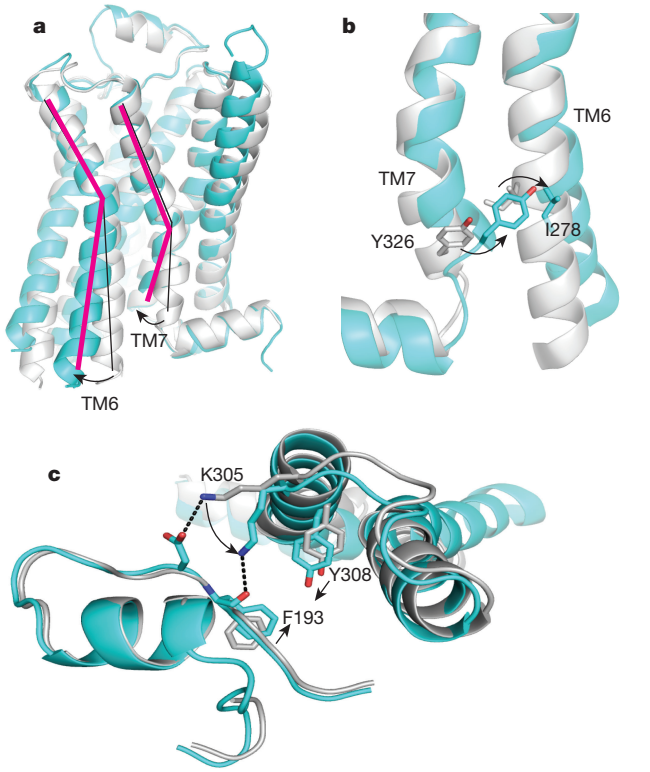


Figure 4 | Allosteric communication between the β_2AR G-protein- and hormone-binding sites. **a**, In the β_2AR active state (cyan), the cytoplasmic end of TM6 moves away from the receptor core by ~14 Å relative to its position in the inactive-state structure, allowing for an inward movement of TM7. **b**, Rotation of TM7 allows Tyr326$^{7.53}$ (of the highly conserved NPxxY motif) to fill the space vacated by the conserved aliphatic residue Ile278$^{6.40}$. **c**, The rotation of TM7 repositions Tyr308$^{7.35}$ and Lys305$^{7.32}$. This conformational change allows Lys305$^{7.32}$ to coordinate the backbone carbonyl of Phe193^{ECL2}, stabilizing its movement towards Tyr308$^{7.35}$ to form a lid over the hormone-binding site.

and Supplementary Video 1). The binding of Gs or Nb80 to the β_2AR stabilizes a rearrangement of the cytoplasmic end of transmembrane domain 7 (TM7; Fig. 4a, b) that is accompanied by changes immediately above the ligand-binding site, as well as a change in the structure of the extracellular loop (ECL) between TM4 and TM5 (ECL2). In comparison to the inactive β_2AR, the structure of the β_2AR–Gs or β_2AR–Nb80 (or related Nb6B9)[14] complex identifies two aromatic residues, Phe193$^{(5.25\text{ or }ECL2)}$ and Tyr308$^{7.35}$, that move approximately 2–2.5 Å closer to each other to form a lid-like structure over the orthosteric ligand-binding site. Lys305$^{7.32}$ also contributes to capping the orthosteric site by trading its salt bridge[15] with Asp192^{ECL2} for an interaction with the backbone carbonyl of Phe193^{ECL2} (Fig. 4c). These structural changes are stabilized in the active forms of β_2AR bound to either the ultra-high-affinity agonist BI-167107 or the smaller, low-affinity agonist adrenaline[15], and formation of this 'lid' would be expected to sterically obstruct both ligand association and dissociation.

To validate this structural model, we tested whether a residue smaller than tyrosine could modify the capacity of Nb80 to slow ligand association. Mutation of Tyr308$^{7.35}$ to alanine, previously shown to lower agonist affinity for the β_2AR[16], significantly diminishes the capacity of Nb80 to slow the association of [^{3}H]DHAP and even the agonist [^{3}H]formoterol (Extended Data Fig. 5), as suggested by recent molecular dynamics simulations[17]. Interestingly, and in contrast to [^{3}H]DHAP association, pre-incubation with 10 μM Nb80 also enhances the extent of [^{3}H]formoterol binding in the Y308A mutant. Eliminating barriers that impair access to the orthosteric site (for example, Y308A) allows the agonist to at least enter the receptor, where it can stabilize nanobody binding. The enhancement, therefore, is a reflection of the capacity of the agonist [^{3}H]formoterol to cooperatively stabilize Nb80 binding and vice versa, and concomitantly slow the dissociation of the bound agonist (Extended Data Fig. 5d). The data also suggest that while Tyr308$^{7.35}$ markedly limits access to the orthosteric site, other residues may work in concert with Tyr308$^{7.35}$ in the active β_2AR conformation to slow agonist dissociation.

It is noteworthy that the movement of Phe193^{ECL2} and Tyr308$^{7.35}$ is not fully observed in the crystal structure of the β_2AR bound to an agonist alone[18], nor in the inactive-state structure of the β_1AR bound to the agonist isoprenaline[19] (Extended Data Fig. 6 and Supplementary Videos 1, 2). Binding of G protein or G-protein mimetic (nanobody) is sufficient to stabilize the closed, active conformation since their effects on ligand-binding kinetics (as in Figs 1 and 2) are agonist-independent. An agonist may enhance G-protein engagement but poorly stabilizes the closed, active conformation by itself. Additionally, the data presented here suggest that formation of the closed, active conformation stabilized by the nucleotide-free G protein can occur owing to basal receptor activity, in keeping with predictions of more recent models of GPCR pharmacology such as the extended and cubic ternary complex models[20,21] (see Supplementary Discussion). Moreover, conformational changes stabilized by the nucleotide-free G protein influence not only agonist binding, but ligand binding in general, implying that the role of nucleotides needs to be included in an updated version of ternary complex model.

© 2016 Macmillan Publishers Limited. All rights reserved

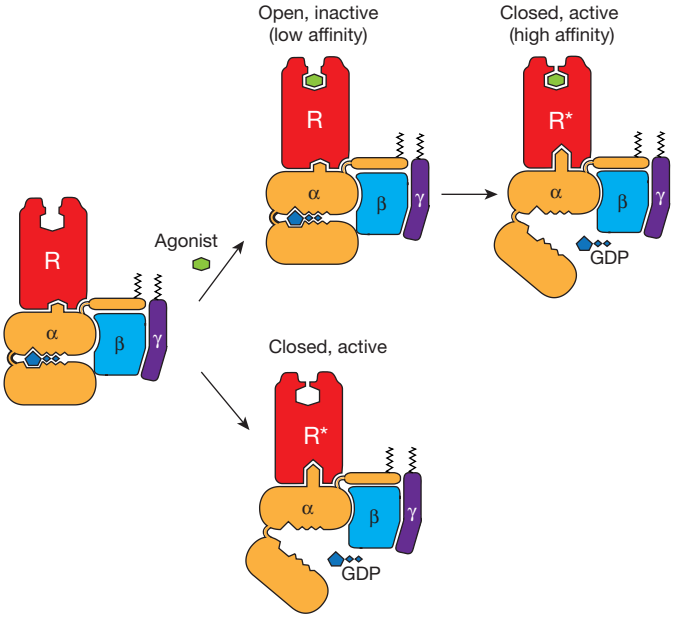


Figure 5 | Basis for G-protein-dependent high-affinity agonist binding. Agonist binding promotes the G-protein–receptor (R) interaction and GDP release from the G-protein heterotrimer (Gα (α) Gβγ (βγ)). In this nucleotide-free state, the C-terminal helix of Gα remains embedded in the receptor core, stabilizing the conformational changes at both the intracellular and extracellular faces of the receptor. At the extracellular side, the orthosteric ligand-binding site closes around the bound agonist, sterically opposing agonist dissociation and thereby enhancing the observed affinity. Constitutive (basal) receptor activity may also activate the G protein, releasing GDP and thereby stabilizing the closed conformation of the receptor in the absence of an agonist. See also Extended Data Fig. 10.

The capacity of G proteins to stabilize a closed receptor conformation explains the poorly defined GTPγS-mediated increase in radiolabelled antagonist binding observed with several GPCRs, including muscarinic, α-adrenergic, adenosine and opioid receptors[22–25] (as in Extended Data Figs 7 and 8). We analysed the behaviour of the M2 muscarinic acetylcholine receptor (M2R) and the μ-opioid receptor (MOPr) to determine whether GTPγS-mediated uncoupling relieves a G-protein-stabilized closed conformation. We focused on these receptors since structural models are available for both inactive and active conformations[26–29], and to determine whether the mechanism we propose for the β_2AR is shared among other GPCRs. The active-state structure of the M2R, in particular, revealed similar conformational changes to the β_2AR in that a lid-like structure is formed above the orthosteric site[27] (see Supplementary Videos 4 and 5). Although the structural changes are not identical, the effect of G proteins (or nanobodies) on the association and dissociation of ligands at the orthosteric sites is shared among the β_2AR, M2R and MOPr (Extended Data Fig. 9 and Supplementary Video 3), suggesting that the allosteric effects of G proteins on orthosteric agonists may be manifested by conceptually common mechanisms. More discussion of the details and implications can be found in Supplementary Discussion. Additionally, many recent studies have focused on the allosteric effect of sodium ions on class A GPCR ligand binding and signalling[30]. Outward movement of TM6 during receptor activation collapses the sodium-binding pocket in many class A GPCRs, thus it appears that loss of bound sodium is necessary for G proteins to stabilize a closed, active receptor conformation.

The formation of the closed conformation is also of particular interest for the development of allosteric modulators targeting class A GPCRs. Most allosteric-modulator-binding sites have focused on the extracellular vestibule located above the orthosteric ligand-binding sites. For the muscarinic M2R for example, the potent M2R allosteric positive modulator LY2119620 utilizes residues that form the lid in the active, closed conformation as described here, as the floor of the vestibule[27]. Stabilization of this closed conformation may therefore be an important aspect on the differentiation between positive allosteric modulators, which enhance agonist binding and activation, and negative allosteric modulators, which decrease agonist binding.

We provide pharmacological and biochemical evidence suggesting that the closed, active conformation of GPCRs is stabilized by the nucleotide-free G protein, allowing G proteins to influence passage of ligands to the orthosteric-binding site. The dramatic effect of G proteins on either ligand association or dissociation is consistent with, and in fact validates, structural models generated from X-ray crystallography in which G-protein coupling on the intracellular face of the receptor allosterically influences the structure of the extracellular face. Agonist or hormone binding enhances G-protein engagement, whereby formation of the active receptor conformation is accompanied by nucleotide loss from the G protein. Therefore, the capacity of G proteins to enhance agonist-binding affinity is structurally and energetically linked to the agonist's capacity to promote nucleotide loss from Gα.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 August 2015; accepted 13 May 2016.
Published online 29 June 2016.

1. Pierce, K. L., Premont, R. T. & Lefkowitz, R. J. Seven-transmembrane receptors. *Nature Rev. Mol. Cell Biol.* **3,** 639–650 (2002).
2. Rasmussen, S. G. *et al.* Crystal structure of the β_2 adrenergic receptor–Gs protein complex. *Nature* **477,** 549–555 (2011).
3. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).
4. Sprang, S. R. G protein mechanisms: insights from structural analysis. *Annu. Rev. Biochem.* **66,** 639–678 (1997).
5. Chung, K. Y. *et al.* Conformational changes in the G protein Gs induced by the β_2 adrenergic receptor. *Nature* **477,** 611–615 (2011).
6. Westfield, G. H. *et al.* Structural flexibility of the Gαs α-helical domain in the β_2-adrenoceptor Gs complex. *Proc. Natl Acad. Sci. USA* **108,** 16086–16091 (2011).
7. Maguire, M. E., Van Arsdale, P. M. & Gilman, A. G. An agonist-specific effect of guanine nucleotides on binding to the beta adrenergic receptor. *Mol. Pharmacol.* **12,** 335–339 (1976).
8. Ross, E. M., Maguire, M. E., Sturgill, T. W., Biltonen, R. L. & Gilman, A. G. Relationship between the β-adrenergic receptor and adenylate cyclase. *J. Biol. Chem.* **252,** 5761–5775 (1977).
9. De Lean, A., Stadel, J. M. & Lefkowitz, R. J. A ternary complex model explains the agonist-specific binding properties of the adenylate cyclase-coupled β-adrenergic receptor. *J. Biol. Chem.* **255,** 7108–7117 (1980).
10. Yao, X. J. *et al.* The effect of ligand efficacy on the formation and stability of a GPCR–G protein complex. *Proc. Natl Acad. Sci. USA* **106,** 9501–9506 (2009).
11. Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469,** 175–180 (2011).
12. Irannejad, R. *et al.* Conformational biosensors reveal GPCR signalling from endosomes. *Nature* **495,** 534–538 (2013).
13. Lefkowitz, R. J. & Williams, L. T. Catecholamine binding to the β-adrenergic receptor. *Proc. Natl Acad. Sci. USA* **74,** 515–519 (1977).
14. Ring, A. M. *et al.* Adrenaline-activated structure of β_2-adrenoceptor stabilized by an engineered nanobody. *Nature* **502,** 575–579 (2013).
15. Bokoch, M. P. *et al.* Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* **463,** 108–112 (2010).
16. Kikkawa, H., Isogaya, M., Nagao, T. & Kurose, H. The role of the seventh transmembrane region in high affinity binding of a β_2-selective agonist TA-2005. *Mol. Pharmacol.* **53,** 128–134 (1998).
17. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **108,** 13118–13123 (2011).
18. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist–β_2 adrenoceptor complex. *Nature* **469,** 236–240 (2011).
19. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β_1-adrenergic receptor. *Nature* **469,** 241–244 (2011).
20. Samama, P., Cotecchia, S., Costa, T. & Lefkowitz, R. J. A mutation-induced activated state of the β_2-adrenergic receptor. Extending the ternary complex model. *J. Biol. Chem.* **268,** 4625–4636 (1993).
21. Weiss, J. M., Morgan, P. H., Lutz, M. W. & Kenakin, T. P. The cubic ternary complex receptor-occupancy model. I. Model description. *J. Theor. Biol.* **178,** 151–167 (1996).

© 2016 Macmillan Publishers Limited. All rights reserved

22. Burgisser, E., De Lean, A. & Lefkowitz, R. J. Reciprocal modulation of agonist and antagonist binding to muscarinic cholinergic receptor by guanine nucleotide. *Proc. Natl Acad. Sci. USA* **79,** 1732–1736 (1982).
23. Bylund, D. B., Gerety, M. E., Happe, H. K. & Murrin, L. C. A robust GTP-induced shift in α_2-adrenoceptor agonist affinity in tissue sections from rat brain. *J. Neurosci. Methods* **105,** 159–166 (2001).
24. Prater, M. R., Taylor, H., Munshi, R. & Linden, J. Indirect effect of guanine nucleotides on antagonist binding to A1 adenosine receptors: occupation of cryptic binding sites by endogenous vesicular adenosine. *Mol. Pharmacol.* **42,** 765–772 (1992).
25. Werling, L. L., Puttfarcken, P. S. & Cox, B. M. Multiple agonist-affinity states of opioid receptors: regulation of binding by guanyl nucleotides in guinea pig cortical, NG108-15, and 7315c cell membranes. *Mol. Pharmacol.* **33,** 423–431 (1988).
26. Haga, K. *et al.* Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **482,** 547–551 (2012).
27. Kruse, A. C. *et al.* Activation and allosteric modulation of a muscarinic acetylcholine receptor. *Nature* **504,** 101–106 (2013).
28. Manglik, A. *et al.* Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature* **485,** 321–326 (2012).
29. Huang, W. *et al.* Structural insights into μ-opioid receptor activation. *Nature* **524,** 315–321 (2015).
30. Katritch, V. *et al.* Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39,** 233–244 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. S. Kobilka for preparation of affinity chromatography reagents and F. S. Thian for help with cell culture. We thank J. Traynor and J. Tesmer for their support and use of their laboratory space for J.P.M. This work was supported by the Lundbeck Foundation (Junior Group Leader Fellowship to S.G.F.R.); Fund for Scientific Research of Flanders (FWO-Vlaanderen) and the Institute for the encouragement of Scientific Research and Innovation of Brussels (ISRIB) (E.P. and J.S.); National Institute of Neural Disorders and Stroke grant RO1-NS28471 (B.K.K.); the Mather Charitable Foundation (B.K.K.); National Institute of General Medical Sciences grants RO1-GM083118 and U19-GM106990 (B.K.K. and R.K.S.) and RO1-GM068603 (R.K.S.); National Institutes of Drug Abuse R21-031418 (B.K.K. and R.K.S.); Michigan Diabetes Research and Training Center Grant, National Institute of Diabetes and Digestive and Kidney Diseases, P60DK-20572 (R.K.S.); University of Michigan Biological Sciences Scholars Program (R.K.S.) and the Rackham School of Graduate Studies (B.T.D.); Molecular Biophysics Training Grant T32GM008270 (B.T.D.); Cell and Molecular Biology Training Grant T32GM007315 (G.A.V.-R.) and Pharmacological Sciences Training Program T32GM007767 (J.P.M.); and AHA Midwest Affiliate Predoctoral Fellowship 13PRE17110027 (J.P.M.).

Author Contributions B.T.D., J.P.M., G.A.V.-R., B.K.K. and R.K.S. designed the experiments. B.T.D., J.P.M., G.A.V.-R. and A.J.K. performed research; S.G.F.R., E.E., J.-J.F., A.M., M.M., Y.D., R.A.M., E.P. and J.S. contributed valuable reagents/analytic tools; B.T.D., J.P.M., G.A.V.-R., B.K.K. and R.K.S. analysed data; and B.T.D., J.P.M., B.K.K. and R.K.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.K.S. (rsunahara@ucsd.edu).

© 2016 Macmillan Publishers Limited. All rights reserved

METHODS

Large-scale purification of the β_2AR. β_2AR bearing an N-terminal Flag tag and C-terminal 10×-His tag was expressed in *Sf*9 cells (Invitrogen) and purified as previously described[2].

Expression and purification of G protein and nanobodies. Gs and Go heterotrimer were expressed in HighFive (Invitrogen) insect cells using recombinant baculovirus and purified by chromatography on Ni-NTA, MonoQ, and Superdex 200 resin, as previously described[31]. Nanobodies were expressed in *Escherichia coli* and purified as previously described[11,14,27].

Membrane preparations. HEK293T cells (ATCC) were used for small-scale expression and purification of β_2AR and mutants. Cells were grown in DMEM plus 10% FBS to ~70% confluency, then transfected with monomeric yellow fluorescent protein (mYFP)–β_2AR (pCMV5, 6 μg DNA per 10-cm plate) using Lipofectamine 2000. Cells were harvested 40–48 h post-transfection in ice-cold lysis buffer buffer (50 mM HEPES, pH 8.0, 65 mM NaCl, 1 mM EDTA, 35 μg ml^{-1} phenylmethylsulfonyl fluoride, 32 μg ml^{-1} each tosyl-L-phenylalanine-chloromethylketone and tosyl-L-lysine-chloromethylketone, 3.2 μg ml^{-1} leupeptin, 3.2 μg ml^{-1} ovomucoid trypsin inhibitor). The cell suspension was sonicated using a Branson Sonifier and centrifuged for 20 min at 25,000*g*. The pellet was resuspended in wash buffer (50 mM HEPES, pH 8.0, 100 mM NaCl with protease inhibitors listed earlier) using a Dounce homogenizer, then centrifuged for 20 min at 25,000*g*. The pellet was resuspended and homogenized in minimal wash buffer and the volume was adjusted to reach a final protein concentration of 5 mg ml^{-1} as measured by the Bradford protein assay. Membranes were frozen by slowly pouring into liquid nitrogen and stored at −80 °C until use.

Enrichment of β_2AR and β_2AR(Y308A) from HEK293T cells. Frozen membranes were thawed on ice and NaCl, $MgCl_2$, and GTPγS were added to reach final concentrations of 300 mM, 1 mM and 10 μM, respectively. Timolol was then added to a final concentration of 1 μM and the membranes were incubated for 10 min on ice. Receptors were solubilized for 1 h at 4 °C in the presence of 1% dodecylmaltoside (DDM) and 0.1% cholesterol hemisuccinate (CHS). After centrifugation for 30 min at 25,000*g*, the supernatant was applied to Ni-NTA agarose. The column was slowly washed with 20 column volumes of 20 mM HEPES, pH 8.0, 300 mM NaCl, 0.1% DDM, 0.01% CHS to remove bound timolol. Receptor was eluted in the same buffer plus 200 mM imidazole and concentrated using an Amicon 30 kDa cut-off spin concentrator for addition to the reconstituted high-density lipoprotein particles (rHDL) reconstitution mixture.

Receptor reconstitution into rHDL particles. Reconstitutions were performed as described[32], with the amount of receptor added never exceeding 20% of the total reaction volume. For samples that contained Gs, the purified heterotrimer was added to the preformed β_2AR–rHDL particles, incubated for 2 h at 4 °C, and BioBeads (Bio-Rad) were used to remove the added detergent. Nucleotide-free Gs•β_2AR complex was prepared by incubating β_2AR–Gs–rHDL particles with apyrase in the presence of 1 mM $MgCl_2$ for 30 min at room temperature, or alternatively, 2 h at 4 °C. If needed, the sample was passed through a Superdex 200 gel filtration column to remove free nucleotide and apyrase.

Radioligand association experiments using rHDL particles. All assays were performed in Tris-buffered saline (TBS; 25 mM Tris-HCl, pH 7.4, 136 mM NaCl, 2.7 mM KCl) with a final concentration of 0.05% w/v bovine serum albumin (BSA). Reaction components were mixed and pre-incubated at room temperature (see later) before the addition of radioligand to initiate the association time course. Aliquots were withdrawn at the indicated times and filtered over Whatman GF/B filters pre-soaked in 0.3% w/v polyethyleneimine. Filters were washed with ice-cold TBS, dried, and subjected to liquid scintillation counting on a TopCount NXT (Perkin-Elmer). Bound ligand never exceeded 10% of the total ligand added.

Kinetic binding experiments with [^{3}H]DHAP and Nb80, β_2AR–rHDL. For association experiments, receptor in rHDL was pre-incubated with varying concentrations of Nb80 and the reaction was started by addition of 5 nM [^{3}H]DHAP (Perkin Elmer). For dissociation experiments, the samples were first incubated with 5 nM [^{3}H]DHAP for 30 min, followed by incubation with varying Nb80 concentrations for 30 min. The reaction was started by adding 50 μM cold alprenolol. Non-specific binding was determined in the presence of 10 μM (+/−)-propranolol.

Binding experiments with [^{3}H]DHAP and Gs•β_2AR nucleotide-free complexes. For association experiments, gel-filtered samples of apyrase-treated Gs•β_2AR–rHDL particles were incubated with 5 nM [^{3}H]DHAP to bind any receptor that was not complexed with Gs. The experiment was started by adding varying amounts of either GDP or GTPγS. For 'equilibrium' binding experiments, samples were incubated with all the indicated components at room temperature for 90 min before filtration. Non-specific binding was determined in the presence of 10 μM (+/−)-propranolol.

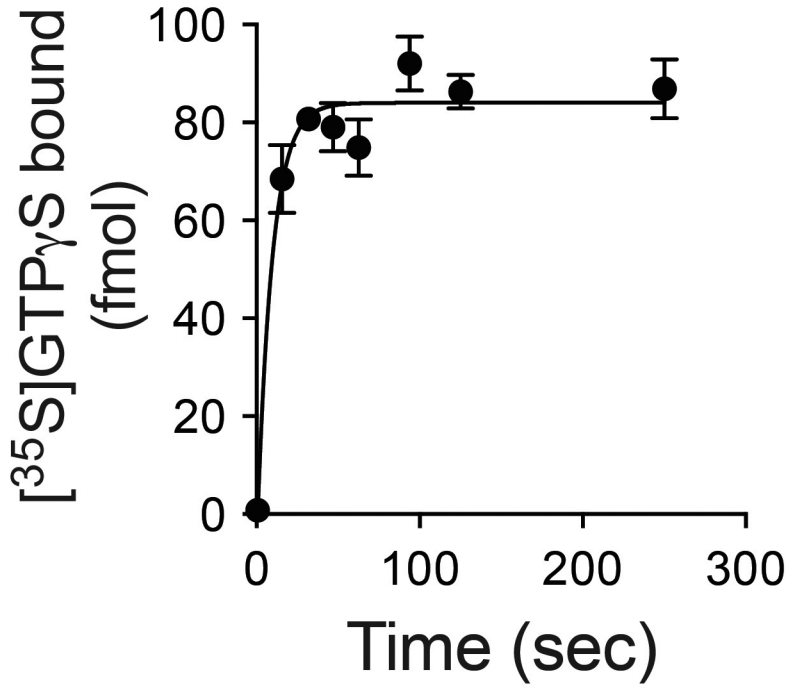
[^{3}H]formoterol association to β_2AR. β_2AR–rHDL was incubated with the indicated concentrations of Nb80 for 30 min at room temperature. [^{3}H]formoterol (Perkin Elmer) was added to reach 10 nM final concentration. These assays also contained 1 mM ascorbic acid in the reaction buffer. Non-specific binding was determined in the presence of 10 μM (+/−)-propranolol.

[^{3}H](−)-CGP-12177 association to β_2AR. β_2AR–rHDL was incubated with the indicated concentrations of Nb80 for 30 min at room temperature. [^{3}H](−)-CGP-12177 (Perkin Elmer) was added to reach 1 nM final concentration. Non-specific binding was determined in the presence of 10 μM (+/−)-propranolol.

[^{3}H]carvedilol association to β_2AR. Owing to high amounts of non-specific [^{3}H]carvedilol (American Radiolabelled Chemicals) binding both to BSA and to the glass fibre filters typically used for separation, β_2AR–rHDL was diluted into empty rHDL particles rather than into a 5× BSA solution (0.25% w/v BSA in TBS buffer) before addition to the assay mix. Using empty rHDL in place of BSA was critical for maintaining sample recovery from the assay plate while improving the signal-to-noise ratio of the assay. The receptor was incubated with the indicated concentrations of Nb80 for 15 min at room temperature, then for 30 min at 4 °C. [^{3}H]carvedilol was added to reach a 1 nM final concentration. Aliquots were withdrawn at the indicated time points and bound ligand was isolated using gel filtration on Sephadex G75 resin. Non-specific binding was determined in the presence of 10 μM (+/−)-propranolol.

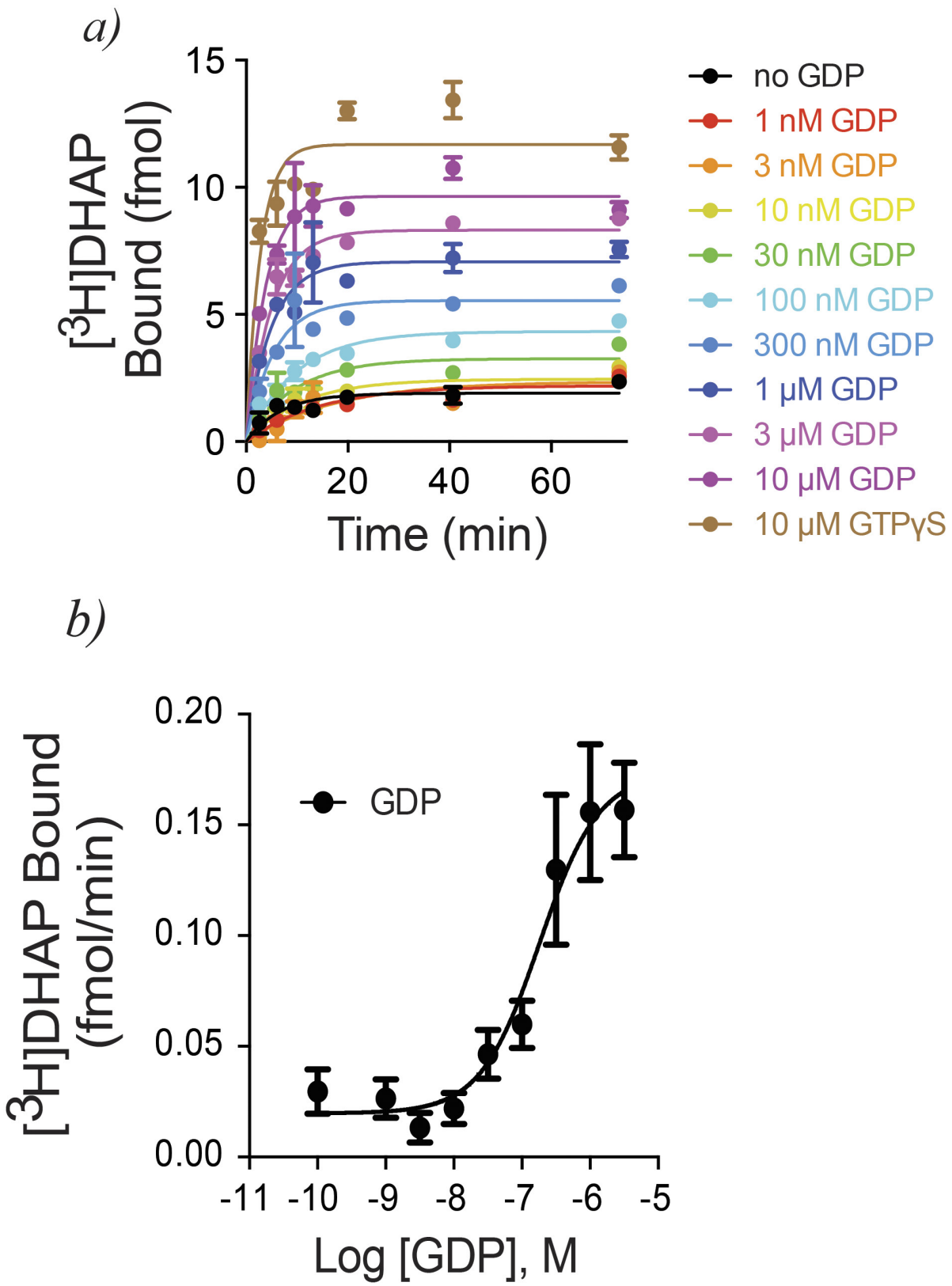
31. Kozasa, T. & Gilman, A. G. Purification of recombinant G proteins from Sf9 cells by hexahistidine tagging of associated subunits. Characterization of α_{12} and inhibition of adenylyl cyclase by α_z. *J. Biol. Chem.* **270,** 1734–1741 (1995).
32. Whorton, M. R. *et al.* A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein. *Proc. Natl Acad. Sci. USA* **104,** 7682–7687 (2007).
33. White, J. F. *et al.* Structure of the agonist-bound neurotensin receptor. *Nature* **490,** 508–513 (2012).
34. Krumm, B. E., White, J. F., Shah, P. & Grisshammer, R. Structural prerequisites for G-protein activation by the neurotensin receptor. *Nature Commun.* **6,** 7895–7895 (2015).

© 2016 Macmillan Publishers Limited. All rights reserved



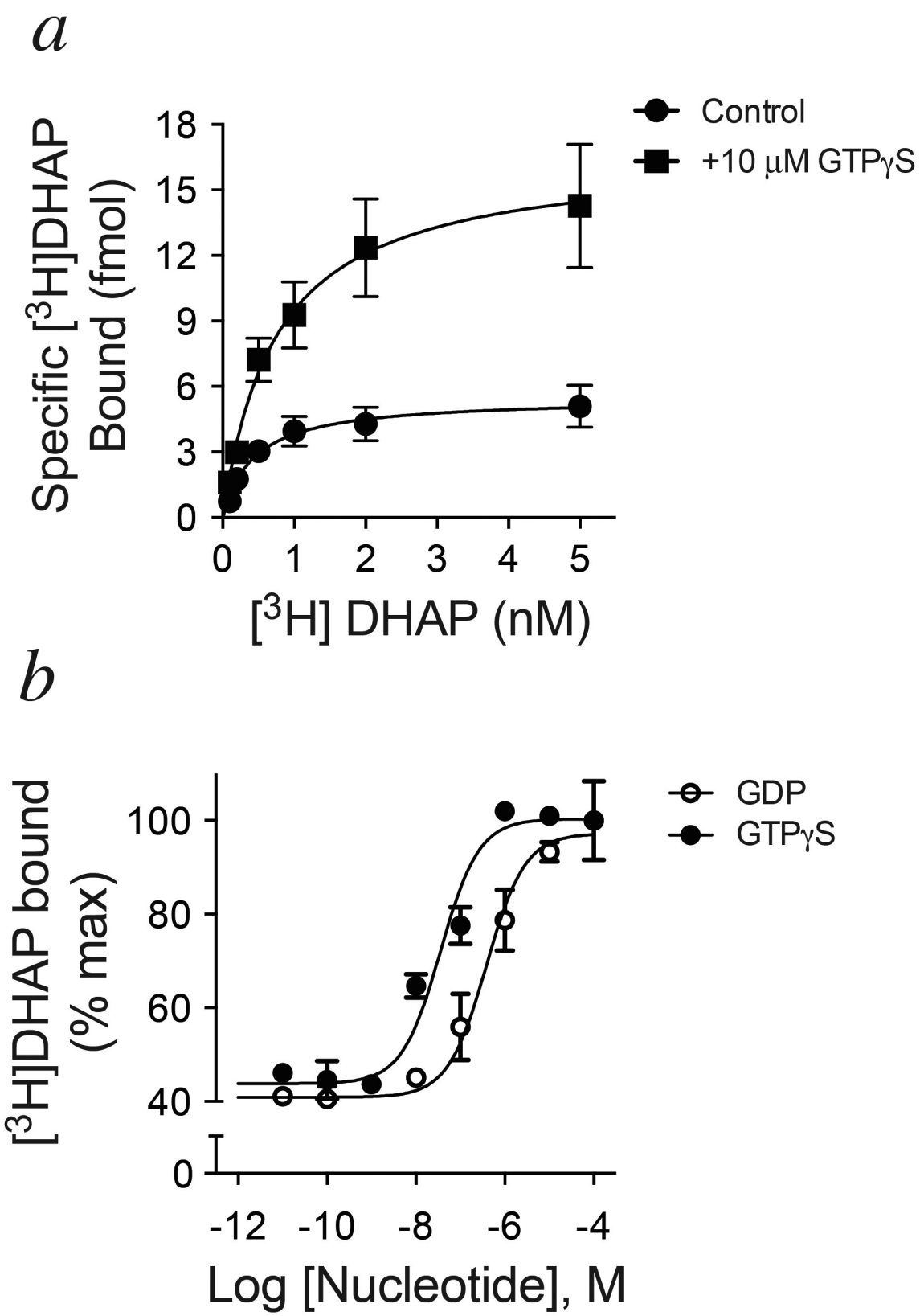
Extended Data Figure 1 | Confirmation of nucleotide removal from β_2AR•Gs by apyrase. Gs and Flag-tagged β_2AR were reconstituted in rHDL and treated with the non-specific nucleotide lyase, apyrase. Samples were applied to an anti-Flag affinity resin to remove products of the GDP degradation (GMP and P_i). Samples were incubated with 100 nM [^{35}S]GTPγS at room temperature. At various times, samples were subjected to rapid filtration through glass fibre filters (GF/B) followed by 10 volumes of ice-cold buffer washes containing 10 μM GDP. Filters were dried and subjected to liquid scintillation counting (Top-Count, Perkin-Elmer). To a first approximation, the rapid binding event suggests that the complex is empty of nucleotide, based on the limited temporal resolution of this mixing and filtration technique. [^{3}H]DHAP and [^{35}S]GTPγS binding to the reconstituted complex yields a final R:G ratio of 1:0.95, suggesting that up to 95% of the β_2AR–rHDL particles contain a single functional G protein. This suggests that only those G proteins associated with the β_2AR will bind [^{35}S]GTPγS within this time frame in the absence of receptor agonists. Data are shown as mean ± s.e.m. from $n = 3$ independent experiments performed in duplicate.

© 2016 Macmillan Publishers Limited. All rights reserved



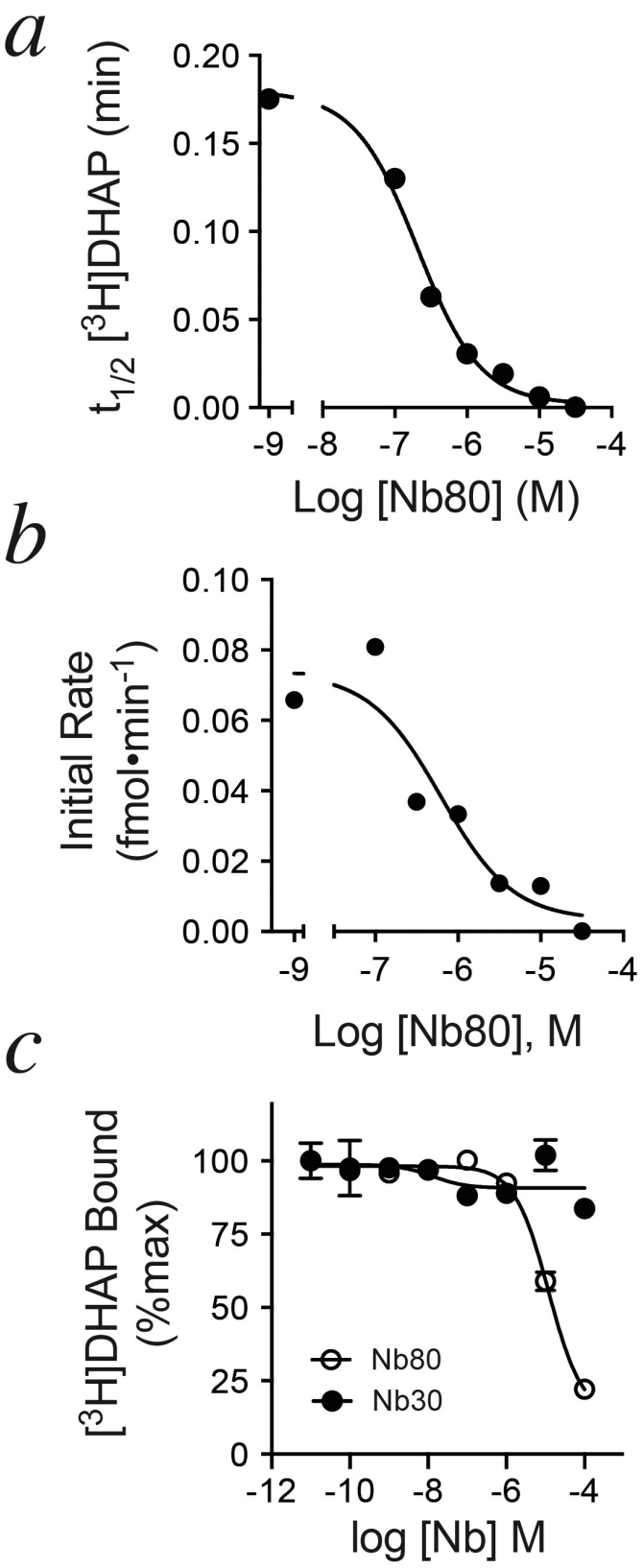
Extended Data Figure 2 | GDP accelerates [^{3}H]DHAP binding to β_2AR•Gs. a, Time course monitoring [^{3}H]DHAP association to apyrase-treated β_2AR•Gs complexes in the presence of varying GDP concentrations. GDP increases both the observed association rate constant and the maximum binding of [^{3}H]DHAP. **b**, Concentration–response curve showing enhancement of the observed [^{3}H]DHAP association rate constant by GDP (half-maximum effective concentration (EC_{50}) = 181 ± 66 nM). All data are shown as mean ± s.e.m. from $n = 3$ independent experiments performed in duplicate.

© 2016 Macmillan Publishers Limited. All rights reserved



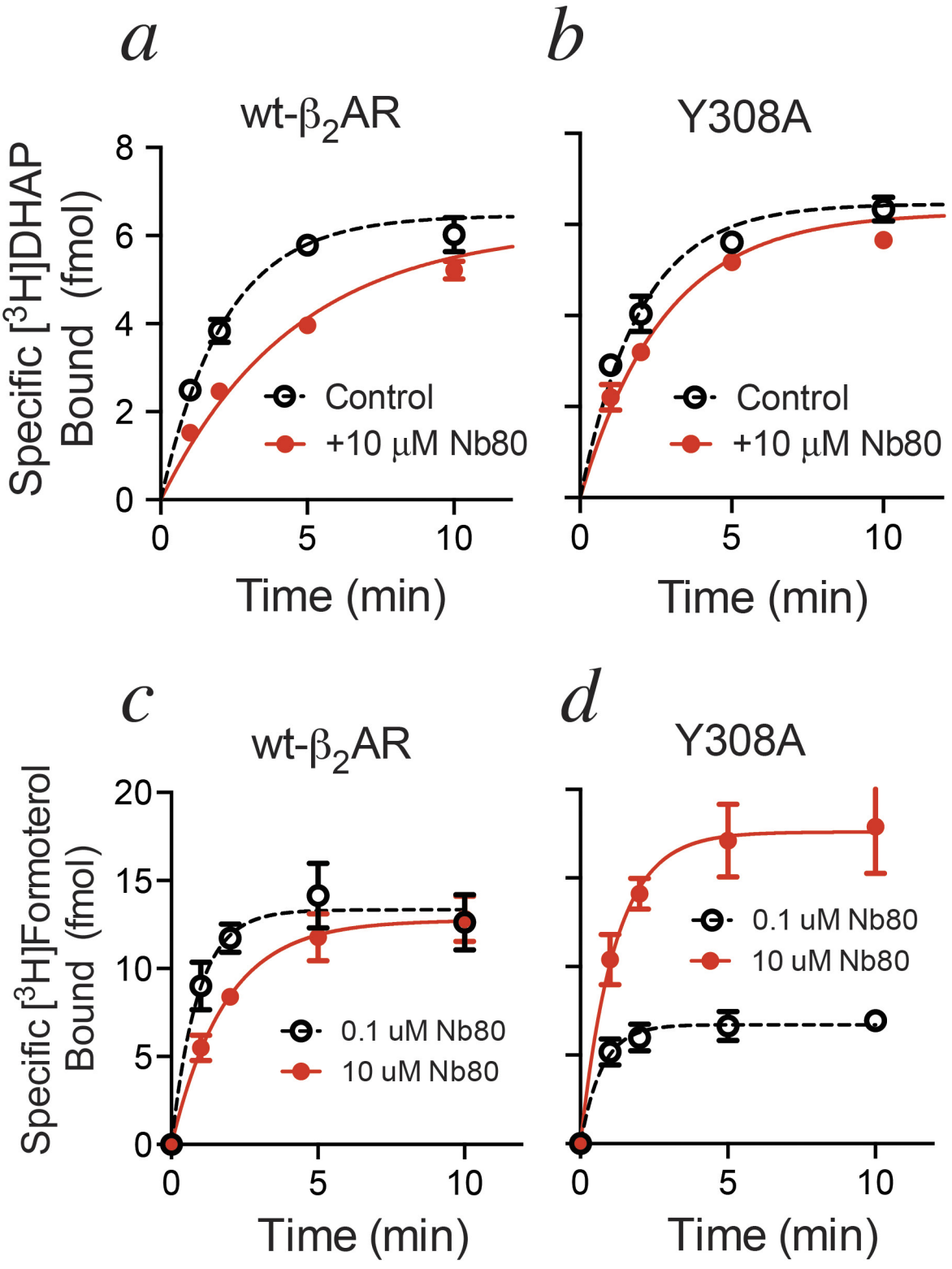
Extended Data Figure 3 | Effect of guanine nucleotides on [^{3}H]DHAP binding to β_2AR•Gs. a, In saturation binding assays, addition of GTPγS to apyrase-treated β_2AR•Gs complexes increased the observed maximal binding, B_{max}, for [^{3}H]DHAP without significantly altering K_d (control: $B_{max} = 5.5 \pm 0.52$ fmol, $K_d = 0.88$ nM; +GTPγS: $B_{max} = 16.6 \pm 1.9$ fmol, $K_d = 0.56$ nM). **b**, Both GDP and GTPγS could enhance maximal [^{3}H]DHAP binding in a concentration-dependent manner (GDP $\log(EC_{50}) = -6.42 \pm 0.12$, or $EC_{50} \approx 386$ nM; GTPγS $\log(EC_{50}) = -7.45 \pm -0.16$, or $EC_{50} \approx 35$ nM). All data are shown as mean ± s.e.m. from $n = 3$ independent experiments performed in duplicate.

© 2016 Macmillan Publishers Limited. All rights reserved



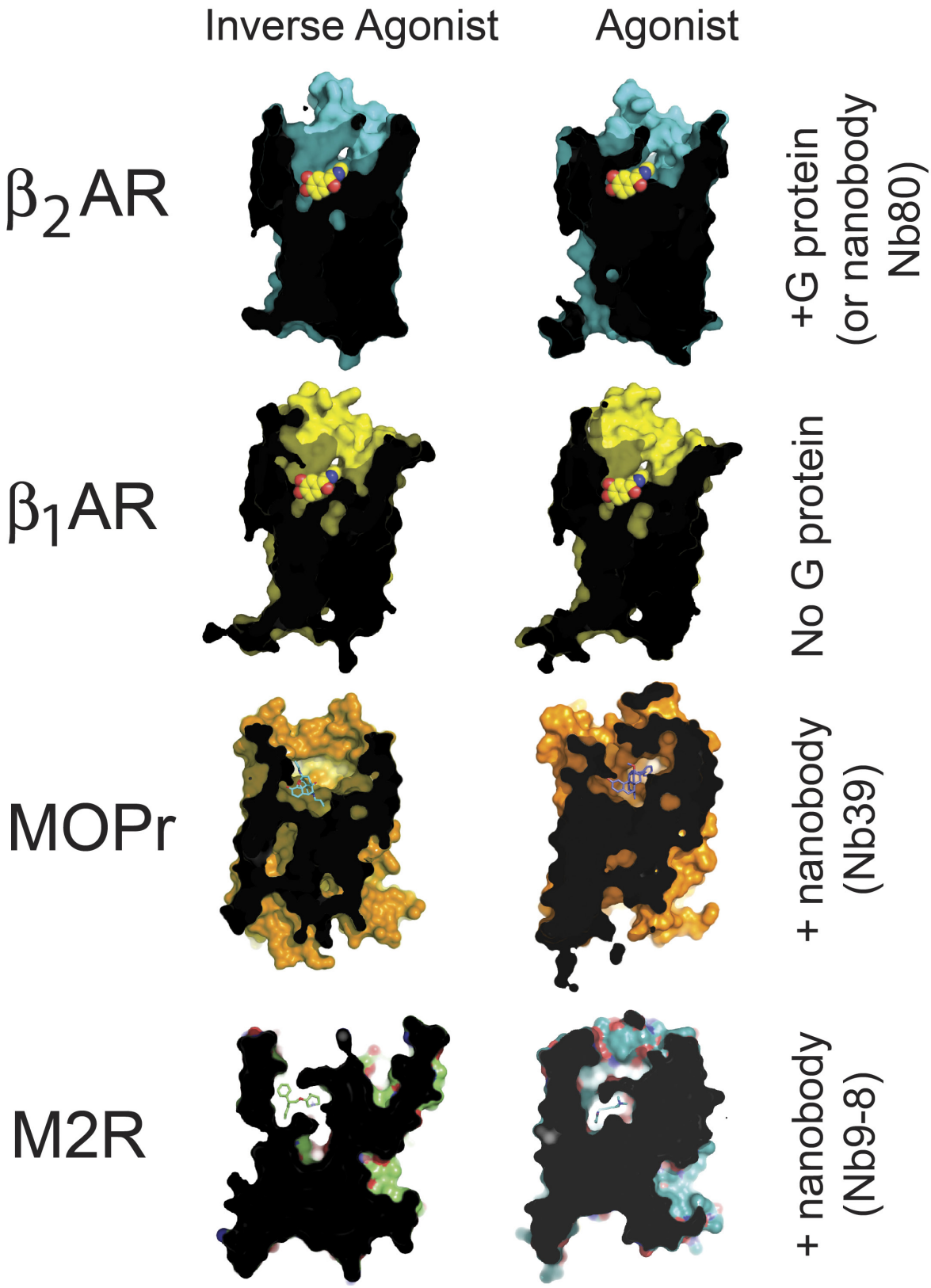
Extended Data Figure 4 | Effect of Nb80 on antagonist binding to the β_2AR. a, Association of [^{3}H]DHAP is progressively slowed after pre-incubation of the β_2AR with increasing concentrations of Nb80. **b**, If [^{3}H]DHAP is allowed to first equilibrate with the β_2AR, Nb80 slows [^{3}H]DHAP dissociation from β_2AR in a concentration-dependent manner. **c**, Owing to the dramatic slowing of [^{3}H]DHAP binding kinetics, Nb80 (but not a control nanobody, Nb30, which has no effect on agonist affinity for β_2AR) seems competitive with [^{3}H]DHAP if insufficient time is given to reach equilibrium. Data shown are from assays incubated for 90 min at room temperature. All data are shown as mean ± s.e.m. from $n = 3$ independent experiments performed in duplicate.

© 2016 Macmillan Publishers Limited. All rights reserved



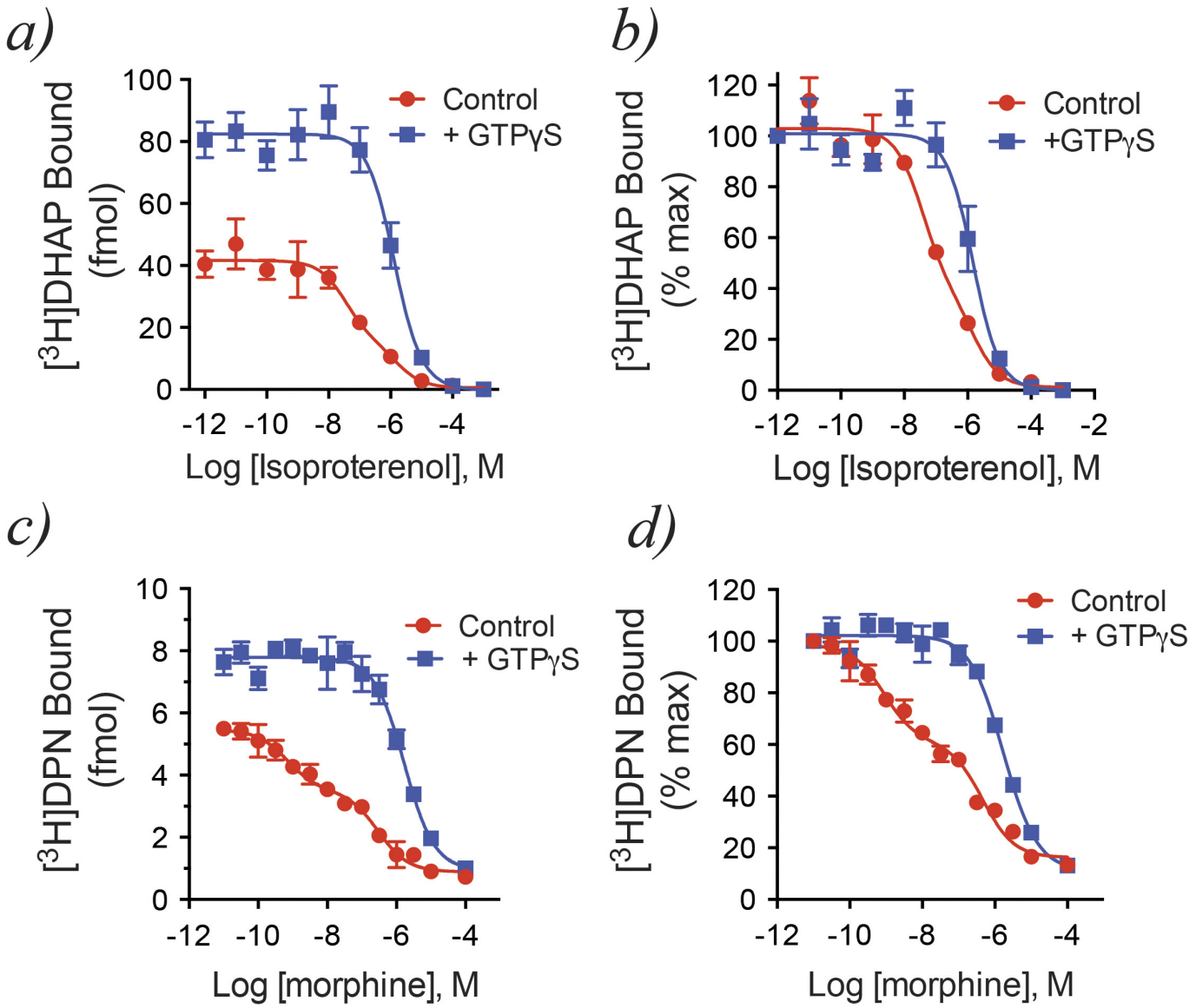
Extended Data Figure 5 | Y308A mutation abolishes the rate-slowing effects of Nb80. a, **b**, Time course of [^{3}H]DHAP binding to wild-type (WT) β_2AR (**a**) or β_2AR(Y308A) (**b**) after pre-incubation of receptor with Nb80. Nb80 significantly slowed [^{3}H]DHAP association to wild-type β_2AR ($-$Nb80 observed rate constant, $k_{obs} = 0.45 \pm 0.05\ min^{-1}$ or association half-time, $t_{½} = 1.5 \pm 0.2$ min, +Nb80 $k_{obs} = 0.20 \pm 0.03\ min^{-1}$ or $t_{½} = 3.5 \pm 0.5$ min; $P = 0.011$ by an unpaired two-tailed t-test), but less effectively slowed [^{3}H]DHAP association to β_2AR(Y308A) ($-$Nb80 $k_{obs} = 0.50 \pm 0.06\ min^{-1}$ or $t_{½} = 1.4 \pm 0.2$ min; +Nb80 $k_{obs} = 0.32 \pm 0.01\ min^{-1}$ or $t_{½} = 2.2 \pm 0.1$ min; $P = 0.05$ by an unpaired two-tailed t-test. All data are shown as mean ± s.e.m. from $n = 4$ ($-$Nb80) or $n = 3$ (+Nb80) independent experiments performed in duplicate. **c**, **d**, Time course of [^{3}H]formoterol binding to wild-type β_2AR (**c**) or β_2AR(Y308A) (**d**) after pre-incubation of receptor with Nb80. Nb80 slowed [^{3}H]formoterol association to wild-type β_2AR (0.1 μM Nb80 $k_{obs} = 0.68 \pm 0.13\ min^{-1}$ or $t_{½} = 1.0 \pm 0.2$ min, 10 μM Nb80 $k_{obs} = 0.27 \pm 0.05\ min^{-1}$ or $t_{½} = 2.6 \pm 0.5$ min; $P = 0.031$ by an unpaired two-tailed t-test). However, with β_2AR(Y308A), Nb80 had little effect on the observed association rate constant but enhanced the amount of [^{3}H]formoterol bound (0.1 μM Nb80 $k_{obs} = 0.37 \pm 0.11\ min^{-1}$ or $t_{½} = 1.9 \pm 0.6$ min with a plateau of 10.1 ± 0.8 fmol, 10 μM Nb80 $k_{obs} = 0.53 \pm 0.13\ min^{-1}$ or $t_{½} = 1.3 \pm 0.4$ min with a plateau of 21.3 ± 1.2 fmol; unpaired two-tailed t-test of the k_{obs} values showed $P = 0.4$). All data are shown as mean ± s.e.m. from $n = 4$ independent experiments performed in duplicate.

© 2016 Macmillan Publishers Limited. All rights reserved



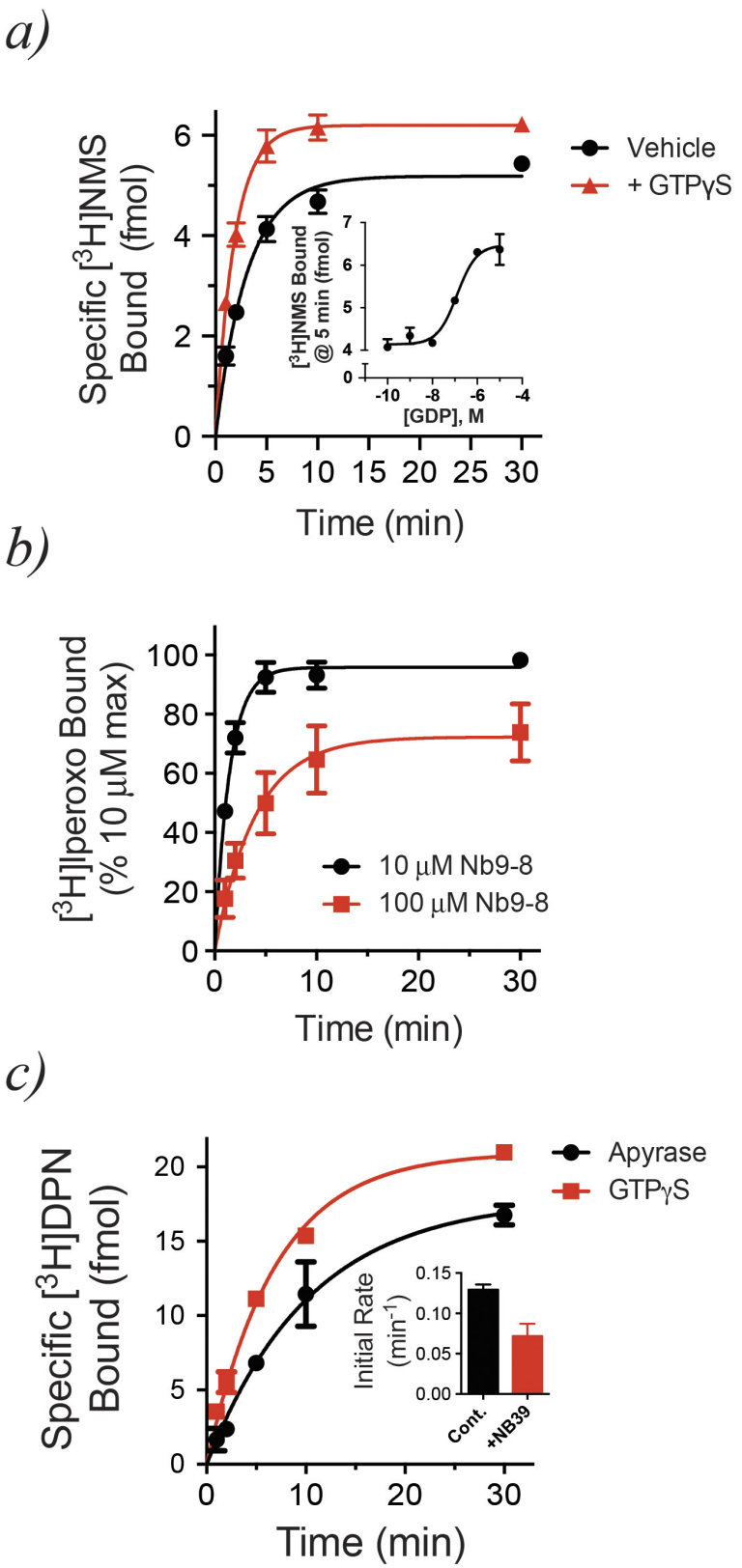
Extended Data Figure 6 | The closed conformation stabilized by agonist and G protein (or mimic). Illustrated are the crystal structures of agonist- versus inverse-agonist-bound β_2AR (cyan) and β_1AR (yellow), where only β_2AR is bound to G protein. Similarly, the MOPr (orange) adopts a closed conformation upon binding the G-protein surrogate, Nb39. β_2AR, PDB accession 2RH1; β_2AR•Gs, PDB accession 3SN6; β_1AR, PDB accession 2YCW; β_1AR-iso, PDB accession 2Y03; MOPr, PDB accession 4DKL; MOPr–Nb39, PDB accession 5C1M; M2R, PDB accession 3UON; M2R–Nb9-8, PDB accession 4MQS.

© 2016 Macmillan Publishers Limited. All rights reserved



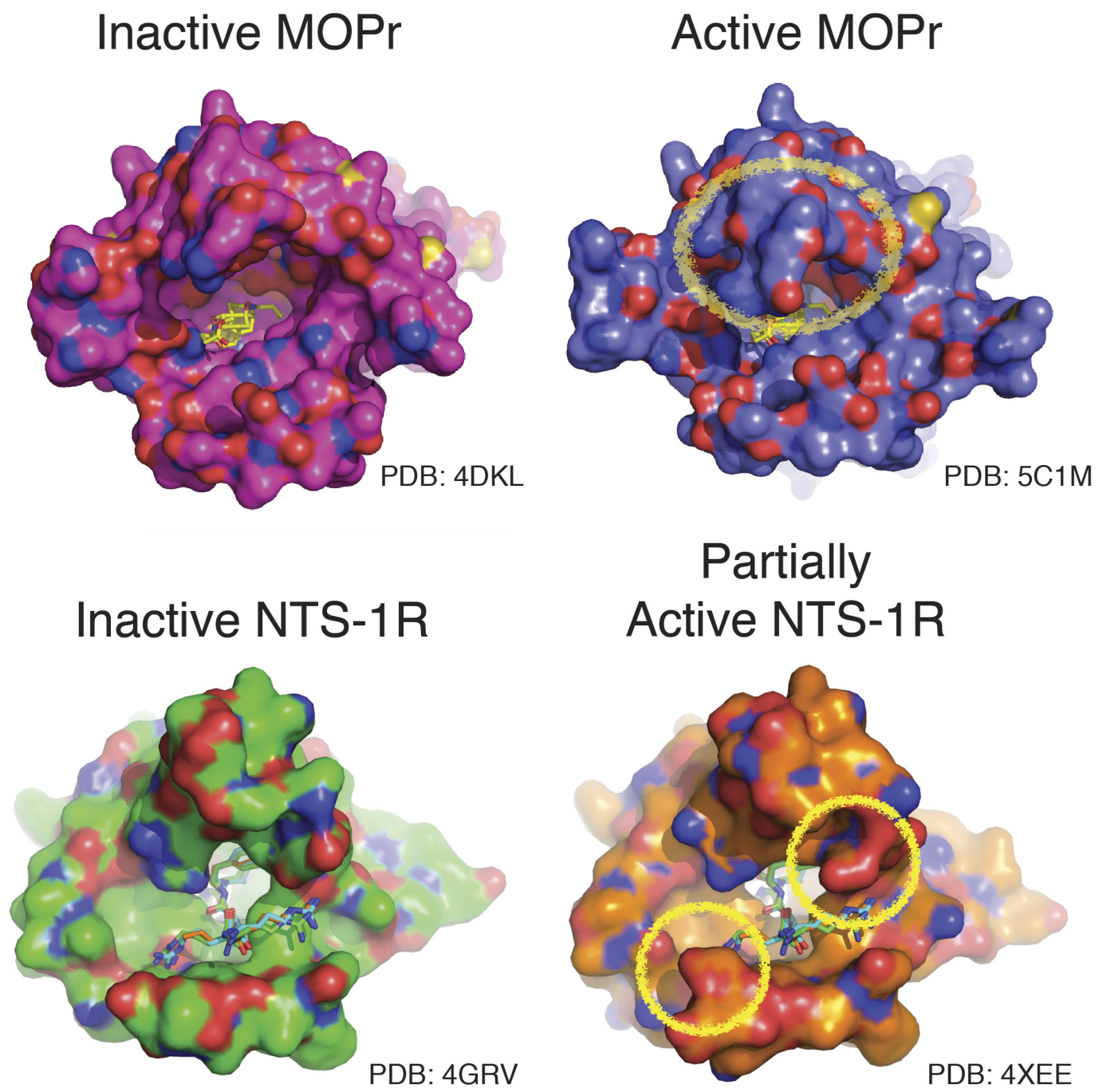
Extended Data Figure 7 | Effect of guanine nucleotides on [^{3}H] antagonist binding are also seen in competition binding assays. **a**, Agonist (isoproterenol) competition binding using apyrase-treated β_2AR•Gs complexes shows the characteristic G-protein-dependent shift in agonist affinity, along with a dramatic increase in total [^{3}H]DHAP binding, upon the addition of 10 μM GTPγS. **b**, Normalization of the data from **a** yields a plot representative of what is commonly reported in the literature. **c**, Similar to the β_2AR, agonist (morphine) competition binding using MOPr•Go complexes shows the characteristic G-protein-dependent shift in agonist affinity, along with a dramatic increase in total [^{3}H]DPN binding, upon the addition of 10 μM GTPγS. **d**, Normalization of the data from **c**. All data are shown as mean ± s.e.m. from $n = 3$ independent experiments performed in duplicate.

© 2016 Macmillan Publishers Limited. All rights reserved



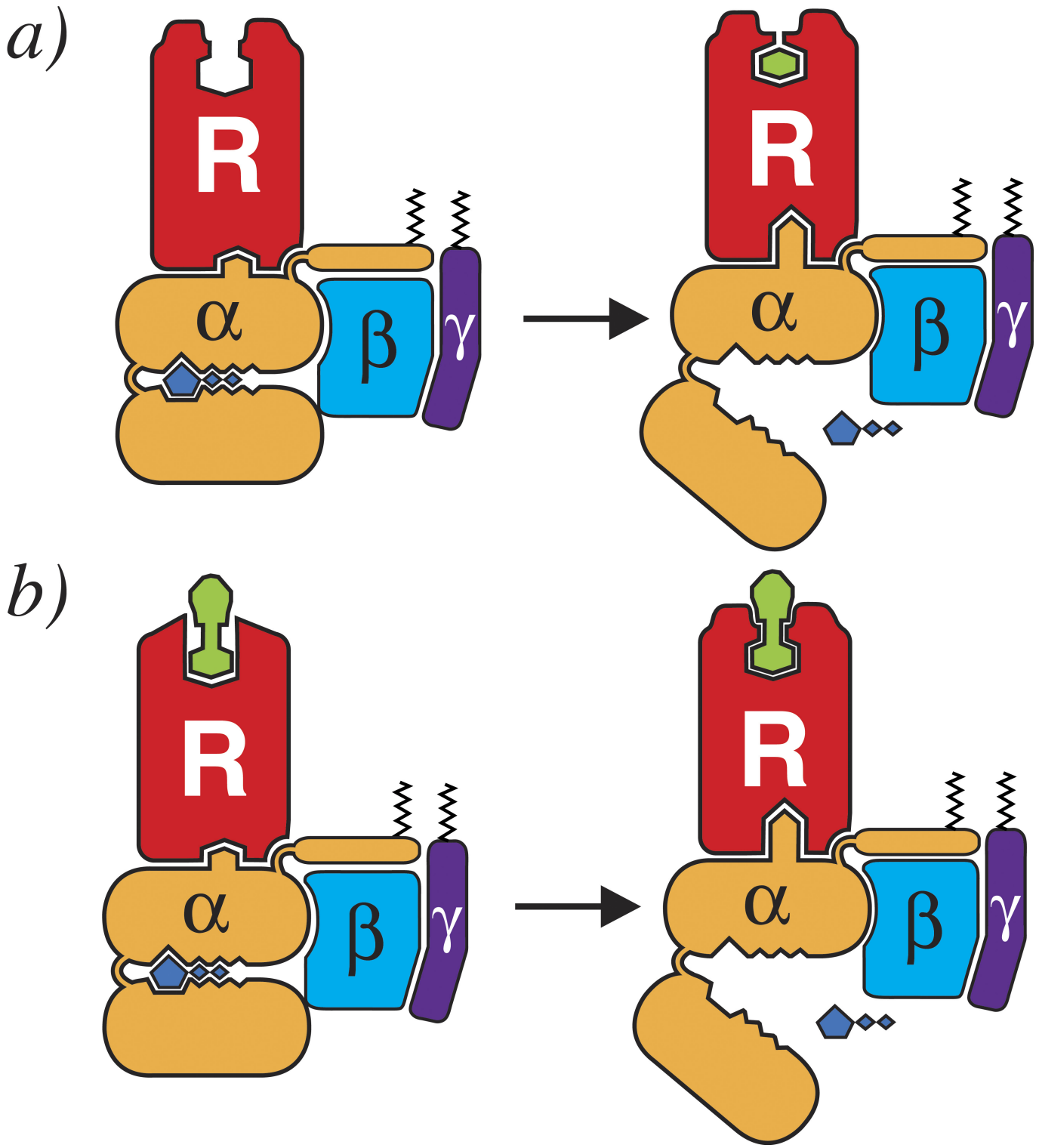
Extended Data Figure 8 | The MOPr and M2R behave similarly to the β_2AR when bound to nucleotide-free G protein or an active-state-stabilizing nanobody. a, After apyrase treatment of M2R•Go complexes, addition of 10 μM GTPγS enhances association of [^{3}H]N-methylscopolamine ([^{3}H]NMS) to M2R (vehicle $k_{obs} = 0.32 \pm 0.02$ min^{-1} or $t½ = 2.2 \pm 0.1$ min, +GTPγS $k_{obs} = 0.54 \pm 0.02$ min^{-1} or $t_{½} = 1.3 \pm 0.1$ min; $P = 0.002$ by an unpaired two-tailed t-test). Data are shown as mean ± s.e.m. from $n = 3$ independent experiments performed in duplicate. Addition of GDP was also able to increase the rate of [^{3}H]NMS binding (inset; log (EC_{50}) = 6.91 ± 0.18 or $EC_{50} \approx 123$ nM; mean ± s.e.m. from $n = 2$ independent experiments performed in duplicate). **b**, Pre-treatment of M2R with either 10 μM (black circles) or 100 μM (red squares) Nb9-8 (ref. 27) impairs association of [^{3}H]iperoxo to M2R (10 μM Nb9-8 $k_{obs} = 0.68 \pm 0.09$ min^{-1} or $t_{½} = 1.0 \pm 0.2$ min, 100 μM Nb9-8 $k_{obs} = 0.25 \pm 0.04$ min^{-1} or $t_{½} = 2.8 \pm 0.5$ min; $P = 0.04$ by an unpaired two-tailed t-test). Data are shown as mean ± s.e.m. from $n = 3$ (10 μM Nb9-8) or $n = 2$ (100 μM Nb9-8) independent experiments performed in duplicate. **c**, Addition of 10 μM GTPγS to apyrase-treated MOPr•Go complexes hastened association of the antagonist [^{3}H]diprenorphine ([^{3}H]DPN) to MOPr (apyrase $k_{obs} = 0.06 \pm 0.02$ min^{-1} or $t_{½} = 9.8 \pm 1.3$ min, +GTPγS $k_{obs} = 0.12 \pm 0.01$ min^{-1} or $t_{½} = 5.6 \pm 0.6$ min; $P = 0.1$ by an unpaired two-tailed t-test). The effect of nucleotide-free G protein was recapitulated by pre-incubating MOPr with Nb39 (ref. 28) (inset; control $k_{obs} = 0.13 \pm 0.01$ min^{-1}, +100 μM Nb39 $k_{obs} = 0.07 \pm 0.02$ min^{-1}). Data are shown as mean ± s.e.m. from $n = 2$ (MOPr•Go) or $n = 3$ (MOPr + Nb39) independent experiments performed in duplicate.

© 2016 Macmillan Publishers Limited. All rights reserved



Extended Data Figure 9 | The extracellular regions in the active conformations of peptide hormone/agonist receptors MOPr and NTS-R1. Illustrated are the crystal structures of the inactive and active (or partially active neurotensin receptor 1, NTS-R1) conformations of the MOPr and NTS-R1 from the top or extracellular view of the receptor. The surface rendering highlights residues or structure on the extracellular face that change upon receptor activation (circled). The MOPr in its inactive conformation (purple) is compared to the Nb39-bound (G-protein mimic) form in blue. Similarly, the inactive NTS-R1 (ref. 33) (green) is compared with a mutant NTS-R1 (ref. 34) that adopts a partially active conformation (orange). MOPr, PDB accession 4DKL; MOPr-Nb39, PDB accession 5C1M; NTS-R1, PDB accession 4GRV; active-like NTS-R1, PDB accession 4XEE.

© 2016 Macmillan Publishers Limited. All rights reserved



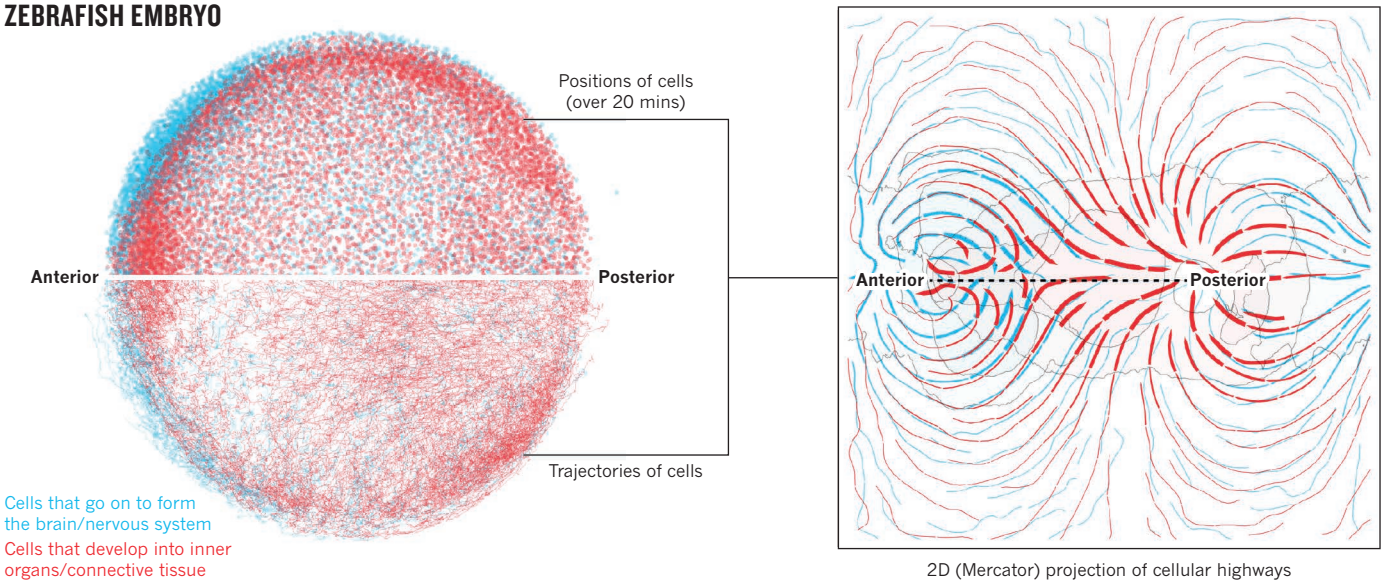
Extended Data Figure 10 | Model of G-protein-dependent high-affinity agonist binding. a, **b**, As in Fig. 5, nucleotide-free G-protein-stabilized family A GPCRs experience alterations in the extracellular face of the receptor, thus affecting the orthosteric-binding site. In a monoamine receptor such as the β_2AR, G-protein binding and GDP loss accompanies the stabilization of a closed, active conformation of the receptor, as in **a**. **b**, For family members such as MOPr or NTS-R1, where the peptide hormones/agonists are considerably larger, the influence of the G-protein-mediated changes in the extracellular domain structure result in similar effects on orthosteric ligand dissociation. Rather than closing over the orthosteric site as with monoamine receptors as in **a**, the extracellular face may contain structures and residues that 'pinch' the larger ligands.

© 2016 Macmillan Publishers Limited. All rights reserved

TOOLBOX

THE VISUALIZATIONS TRANSFORMING BIOLOGY

Inventive graphic design and abstract models are helping researchers to make sense of a glut of data.



NICO SCHERF, KONSTANTIN THIERBACH, GOPI SHAH, INGO ROEDER, JAN HUISKEN

The 'flow' of cells in a developing zebrafish embryo, seen in 3D microscopy data (left) and as a 2D projection (right).

BY EWEN CALLAWAY

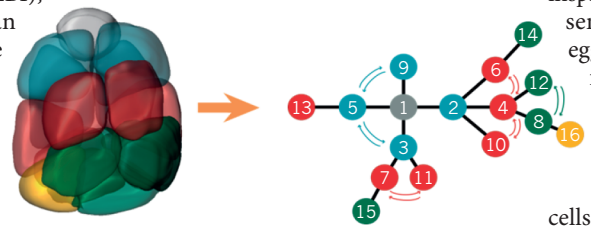
A smart visualization can transform biologists' understanding of their data. And now that it's possible to sequence every RNA molecule in a cell or fill a hard drive in a day with microscopy images, life scientists are increasingly seeking inventive visual ways of making sense of the glut of raw data that they collect.

Some of the visualizations that are currently exciting biologists were presented at a conference at the European Molecular Biology Laboratory in Heidelberg, Germany, in March. Called Visualizing Biological Data (VIZBI), the meeting was co-organized by Seán O'Donoghue, a bioinformatician at the Garvan Institute of Medical Research in Sydney, Australia. The gathering attracts an eclectic mix of lab researchers, computer scientists and designers and is now in its seventh year.

Here, *Nature* highlights some of O'Donoghue's picks of the visualizations that are set to transform biology.

STREAMLINED CELLS

Bioinformatics postdoc Nico Scherf watches cells shift paths to form different germ layers and then organs in developing zebrafish embryos using 'light-sheet microscopy' techniques developed by his supervisor at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany. But, he says, when tracing the path of every single zebrafish cell, "you end up with a hairball" of tracks. To get some meaning out of these hairballs, Scherf borrowed some fluid-dynamics approaches used to analyse atmospheric and ocean currents. "You only plot the major streamlines, which gives you the highways of cellular motion," he says. To achieve this, Scherf wrote some software to analyse the images, and will share it with others on request. So far, his approach has revealed that a mutation that causes abnormal organ development alters the movement of cells only very early in zebrafish development. And he thinks that people who are studying the development of other organisms could benefit from getting into the flow of things, too.



JASMIN IMRAN ALSOUS & PAUL VILLOUTREIX

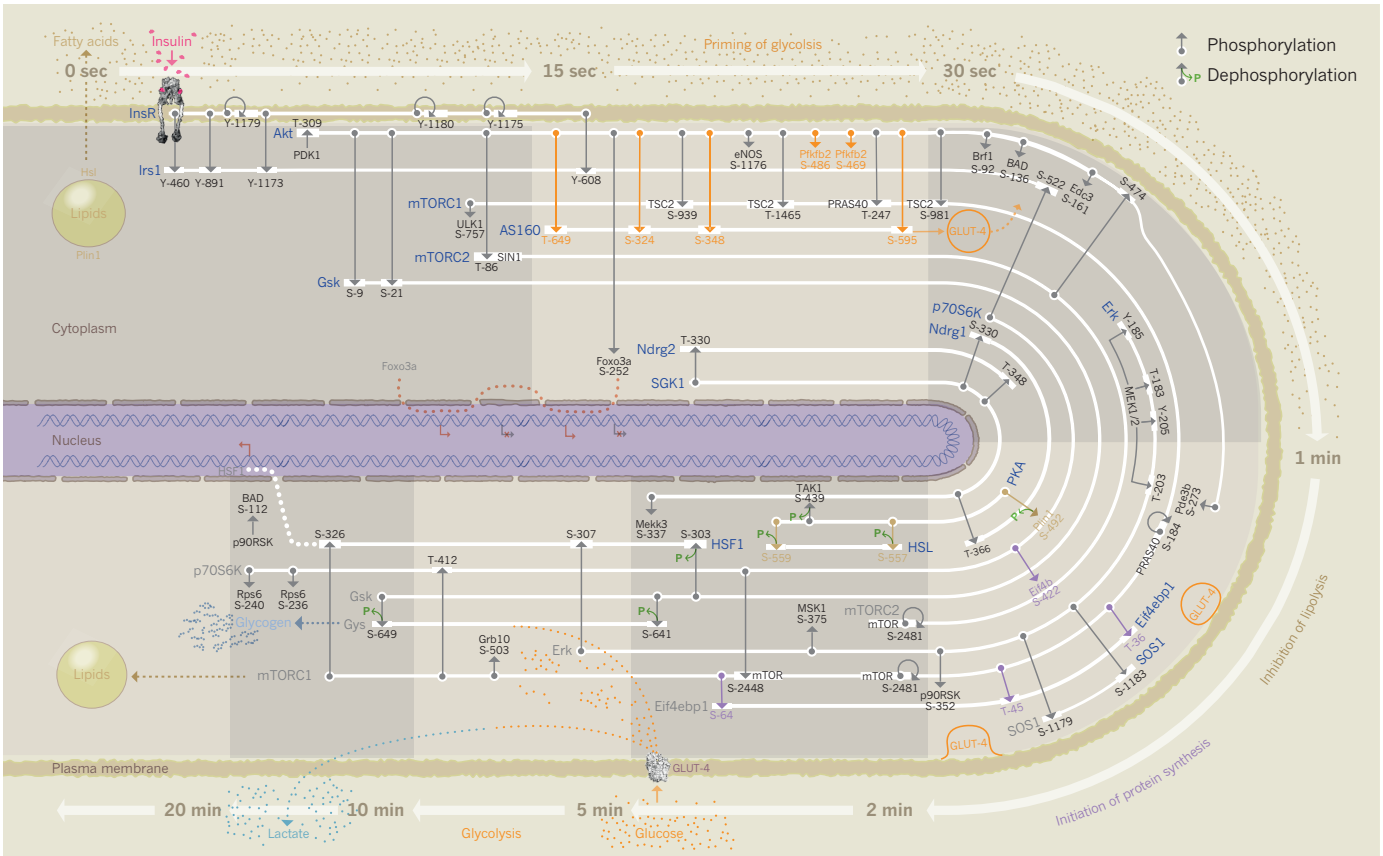
Cell interconnections in a *Drosophila* egg chamber (left) are represented as a 1D network (right).

ABSTRACT CONNECTIONS

Jasmin Imran Alsous, a developmental biologist at Princeton University in New Jersey, took inspiration from Picasso when trying to make sense of microscopy images of a fruit fly's egg chamber, a torpedo-shaped cluster that forms when a germ cell goes through four incomplete and asymmetrical divisions. The final result is a network of 16 interconnected cells that constitute both the developing embryo and the surrounding cells that nourish it.

Alsous's adviser had sent her an article about Picasso lithographs that depicted ▶

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



A 'Minardo' chart that visualizes a cascade of protein phosphorylation after a cell is treated with insulin.

increasingly abstract renderings of a bull. She thought the same principle could apply to depictions of the egg chamber.

She transformed fluorescent microscope images of the chamber into a string of numbers that unambiguously represent how each cell connects to the others. Using this abstraction, she has found that some of the 72 possible configurations of the egg chamber are much more common than others. She is now testing whether the different configurations affect how fruit-fly embryos grow and develop.

A BETTER MODEL OF THE CELL

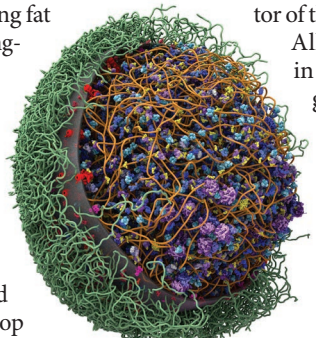
O'Donoghue says that his first attempt to visualize how fat cells respond to insulin ended up as a hairball of criss-crossing molecular pathways. A colleague had measured how hundreds of different protein types in a cell are phosphorylated (which tends to activate them) in response to insulin over the course of an hour, when the cell stops burning fat to produce energy and starts bringing in sugars and storing fats.

To tame the hairball, O'Donoghue found inspiration in a famous chart created by the nineteenth-century French civil engineer Charles Joseph Minard. The image charted Napoleon's disastrous invasion of Russia, and integrated six kinds of data — including troop numbers and geography — in two dimensions. O'Donoghue's 'Minardo' chart visualizes an insulin-treated cell like a clock, with consecutive phosphorylation events moving clockwise around the cell. It also depicts a protein's location in a cell and its relationship to other molecular players.

One of the major insights from the visualization, O'Donoghue says, is how quickly the cell responds to insulin, with many changes occurring in the first 15 seconds. "A lot of people in the community were quite shocked by the suddenness." He is eager for others to use the approach to map other dynamic events, such as the cell cycle, and has created an online guide for doing so. But, for now, he says, "you have to do a lot of manual tweaking".

INSIDE OUT

Illustrator Graham Johnson is used to depicting the internal life of cells by hand. Now director of the Animated Cell project at the Allen Institute for Cell Science in Seattle, Washington, Johnson got his start doing illustrations for a cell-biology textbook. "Despite painstaking efforts to be accurate, it was always easy to make mistakes," he says — particularly when depicting the relative size of cellular components. "When you're creating a visualization, you're establishing what will be the mental model for many current and future biologists," so accuracy is key, he adds.



A CellPACK 3D molecular model of a *Mycoplasma mycoides* cell.

To make cellular model-making more systematic, Johnson developed a tool called cellPACK. To use it, researchers use experimental data to create a series of physical rules (a 'recipe') by which defined cellular components such as proteins, lipids and nucleic acids (the 'ingredients') fill a space. Johnson would like to create a platform such that the models are automatically updated when new data are generated. But despite lots of interest from other researchers, most life scientists find that the tool requires too much time and effort to be very practical. "It's months of research to generate a recipe from scratch," says Johnson, who plans to release a more streamlined web version of the software later this year.

The tool isn't just for making visually striking models, he emphasizes. It can also help scientists to come up with testable hypotheses. His team created a model of the internal structure of HIV and used it to predict how the protein that forms the outer shell interacts with an internal protein. Johnson says that a virologist recently got in touch with him to say his conclusions gleaned from cellPACK checked out experimentally. "He has a bunch of new data, and he wants to work with us to build new models." ■

D. K. G. MA *ET AL. CELL* **161**, 948 (2015)

G. T. JOHNSON *ET AL. NATURE METH.* **12**, 85–91 (2015)

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

CAREERS

GERMAN ACADEMIA One thousand new tenure-track posts **p.190**

PUBLISHING JOBS A production editor describes her day **go.nature.com/296ghpw**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

DARIO PIGNATELLI/BLOOMBERG VIA GETTY



A stint as a postdoc is beneficial whatever research career students are intending to pursue.

COLUMN

Keep it moving

A postdoc job is good for your career, but don't get stuck in an academic cul-de-sac, says **Søren-Peter Olesen**.

Should you take up a postdoctoral position after earning your PhD? My view is that you should. This is provocative advice in the face of data that clearly substantiate a worldwide oversupply of researchers who have completed such a post. Yet I am not suggesting that you undertake multiple postdocs, as many junior researchers do. Instead, I believe that a single postdoc term will benefit your career if you want to stay in research.

As director of the Danish National Research Foundation (DNRF) in Copenhagen, I've watched numerous trainees, including those in my own lab, make remarkable progress within a short time when they are exposed to the challenges that a postdoc role provides. These challenges offer an ideal background for rethinking and redefining your career away from academia (or in it, if you're one of the few fortunates). I've gathered evidence from junior researchers who have worked in my lab, and from interviews with and a survey of former postdocs that further support my advice. Industrial employers from leading research-intensive companies in Denmark, to whom we presented these results, told us that they prefer candidates who have completed one term of postdoctoral research.

A postdoc placement is, of course, almost obligatory for an academic-research career — but the unfortunate and often-cited reality is that few tenure-track posts are available anywhere in academic research.

Yet a postdoc is valuable to you no matter what research career you pursue or in which sector you pursue it. You further develop your scientific and research skills and talents by working more independently on original problems, using innovative techniques; and you complement the abilities that you acquired during your PhD programme. In a postdoc role, you take more responsibility for the research; you learn how to manage others and apply for funds; and you are likely to receive greater exposure to a workplace in which many of your colleagues are from different nations. These are highly useful competencies for a research position in any sector.

Does a single postdoctoral stint help you to win an industrial research position? I believe that it does. I have watched junior researchers in my lab advance smoothly into industrial research careers after one postdoc term. Of the roughly 30 postgraduates who have left my lab over the past 11 years for a research job in industry, 21 obtained such jobs after one postdoc. Many of those successful candidates had competed with up to 100 other applicants.

With a single postdoc behind them, young scientists are highly attractive to the scientific community. The Danish industry representatives who attended our presentation of survey and interview results stated unequivocally that they would rather hire scientists who had completed one postdoc at a highly ranked international university than someone who had just finished a PhD. And they reiterated their stance at a round-table discussion in June. But although one postdoctoral stint provides great value, the same cannot be said for two or more. The same industrial employers said that they might lose interest in candidates who have done many years of postdoctoral training.

To be sure, the glut of researchers who have finished postdocs is no different in ▶

© 2016 Macmillan Publishers Limited. All rights reserved.

Denmark than it is anywhere else. Their numbers, and the number of non-tenured assistant professors, nearly doubled between 2006 and 2013, reaching 3,598, whereas the number of associate professorships grew by less than 25%, to 4,443. Of these, just 5–10% become available each year. Many people with postdocs work hard on short-term contracts while waiting for a vacant professorship. Most will wait in vain.

Many junior scientists do multiple postdocs, in part to further their dream of a professorship and partly because they see no clear alternative. But it is clear from our survey of and interviews with former postdoc researchers (which we conducted between 2014 and 2015) that aiming for academia through multiple postdocs is unlikely to bring career satisfaction. The 400 participants had done postdocs between 2007 and 2014 at DNRF centres of excellence (research units embedded in Danish universities or research institutions). Of the 20% who now work in industry as researchers or managers, 85% said that they were very or fairly satisfied with their current job. And they reported greater satisfaction with their job security and career opportunities than did those in academia, including researchers currently doing a postdoc.

Yet half of the interviewees and survey respondents consider it unlikely or very unlikely that they will get a non-academic job, mainly because they think that they lack the necessary competencies. Most postdoc researchers whom I have interviewed also believe that they are on the path to a career in academia — though the disheartening truth is that even if you are a great scientist, there is often no place for you there. But it is clear from our survey and interviews that many people do up to three postdocs, increasing the risk that a potential employer, especially in industry, will see them as too specialized.

Ask yourself and your supervisor during your first postdoc whether you should aim beyond an academic career — and demand career advice and mentoring from people who work in relevant research-based industries or in the public sector. You need strong and specific career advice, including exposure to role models with careers outside academia. Only 20% of the postdocs in our survey had received such guidance.

You must control your own career. Don't languish in a sector in which there might be no position for you, even if it seems risky to leave academia. A willingness to take risks is characteristic of a great professional life. As the Danish philosopher Søren Kierkegaard said: "To dare is to lose one's footing for a while. Not to dare is to lose oneself." ■

Søren-Peter Olesen *is director of the Danish National Research Foundation in Copenhagen.*

UNIVERSITY JOBS

Germany to fund tenure-track posts

Federal government will create 1,000 professorships.

BY AMBER DANCE

German President Angela Merkel and state prime ministers have signed a €1-billion (US$1.1-billion) agreement to fund 1,000 new tenure-track professorships, in the hopes of retaining and recruiting top academic talent in the nation.

According to the *Nachwuchspakt* ('junior pact'), as the contract is known, the federal government will pay young professors as they work towards tenure, after which state-funded universities will assume financial responsibility.

"It's the first time that the federal government, as far as I know, is investing such a lot of money into the careers of young scientists," says Christian Schäfer of the German Academic Exchange Service in Bonn. The agreement, signed on 16 June, reflects an effort to improve the job situation for young researchers in Germany, where tenure-track positions are rare. Scientists typically work in temporary posts until they are eligible for a faculty spot — usually not until their early 40s, at which point it is difficult to start a non-academic career.

"Because of the perceived insecurity, there are great minds who leave the academic world."

Schäfer and many young researchers say that the agreement is a positive step — but that more needs to be done. "It's better than nothing," says Andreea Scacioc, a structural biologist in Göttingen, who earned a PhD in 2014. "But it's too little."

Every year, about 28,000 PhD and medical students graduate from German universities. There are about 25,000 actively employed professors, according to the German Association of University Professors and Lecturers (DHV). The Society of Junior Professors, a national advocacy group for junior academics, has argued that tenure track ought to be the default entry-level post for junior academics, and DHV officials estimate that 7,500 more professorships are needed to offer young academics a better future.

The pact will run from 2017 to 2032 and involve two major hiring waves, in 2017 and 2019. Universities must apply for funds to set up these professorships. The federal government will fund the first six years of a professor's position, as well as two extra years for those who earn tenure. But researchers will still need to obtain grant funding because the pact funds will mainly cover their salary. Fifteen percent of the total money will be set aside for universities to develop research career paths — for example, by instituting other kinds of permanent positions.

German universities tend to hire few permanent professors. Those who are hired run a 'mini-department', says Jakob Macke, a computational neuroscientist at the Max Planck-affiliated neuroscience-research centre Caesar in Bonn. The general route to independence has been to perform a *Habilitation* — a sort of second thesis — under a professor's guidance, which qualifies a postdoc for a professorship.

Starting in the late 1990s, German institutions introduced various sorts of junior professorships and group-leader positions. These allow young researchers to skip the *Habilitation* and run their own labs, but they are temporary — and many researchers still do a *Habilitation*. "Because of the perceived insecurity, there are great minds who leave the academic world," says Jens Pöppelbuß, a junior professor of industrial services at Germany's University of Bremen. Other talented scientists decamp for nations that offer more direct career paths.

The *Nachwuchspakt* arose in part from changes to Germany's 2005 Excellence Initiative, which funded graduate schools; 'clusters of excellence' that offered international-scale training and research facilities; and competitive research programmes. The original initiative will expire in 2017, and the new version — also signed on 16 June — will drop its focus on graduate schools and early-career scientists, leaving a hole that the Nachwuchspakt will fill.

But Scacioc points out that the pact does not set a quota for hiring women. She fears that it could perpetuate the status quo in which men are more likely to secure professorships, thanks in part to their winning more prestigious awards. Requirements for hiring and tenure will need to be clear and transparent to keep the process fair and to ensure that the best candidates get the positions, says Jule Specht, a personality psychologist at the Free University of Berlin.

"Money from the federal government can only provide some incentives," says Pöppelbuß. "All the different federal states and all universities must commit themselves to establishing more reliable and predictable career paths in academia." ■

© 2016 Macmillan Publishers Limited. All rights reserved.

LIFE IN THE CLOUDS

A long-term problem.

BY DAVID B. LITT

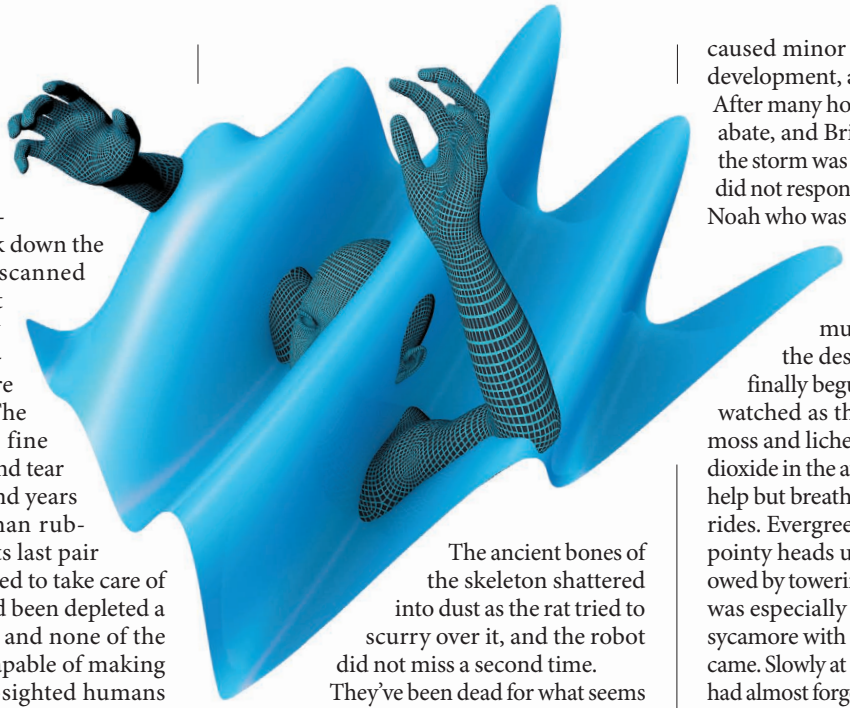


ILLUSTRATION BY JACEY

RH0V3R-4 took aim and fired. A hairless ape gave a high-pitched yelp and fell down. Its companions turned and fled back down the dusty trail. RH0V3R-4 scanned the cloudy skyline as it turned back and slowly rolled towards the subterranean compound where its creators lay in state. The dusty path used to be a fine concrete road, but wear and tear over the past few thousand years had reduced it to less than rubble. RH0V3R-4 was on its last pair of good wheels, and wanted to take care of them. The stockroom had been depleted a few hundred years back, and none of the robots were physically capable of making more. Funny how short-sighted humans were. Causing their own demise, and then not even bothering to make their children capable of reproduction. That was the fatal flaw of species with short life spans. They never thought about long-term problems. When the long-term problems of the past became the immediate ones of the present, the solutions they came up with were unexpected, to say the least.

RH0V3R-4 beeped a hello, and F1X3R-6 acknowledged it by squirting some blue paint in a rastering pattern on the wall, restoring a sign that said: *Heavenly Storage™: Your #1 Site for All Your Cloud Service Needs.* What a waste, thought RH0V3R-4.

The F1X3R class had done a horrible job of upkeep. They were woefully underprepared for fighting mould and termites, but were wonderful at vanity projects — like putting a shiny new case on a blown-out and corroded motherboard.

RH0V3R-4 put its criticisms out of circuit as it emerged in a subterranean concourse that was full of thousands of once sparkling fMRI machines. They were grey with dust. The ceiling was cracked and had been repaired a hundred times, but RH0V3R-4 didn't know why they still bothered. A rat scurried down an aisle between the brain scanners. RH0V3R-4 aimed and fired. The rat dodged it by an inch and leapt onto an fMRI machine. The ancient bones of the skeleton shattered into dust as the rat tried to scurry over it, and the robot did not miss a second time.

They've been dead for what seems like forever, thought RH0V3R-4. They uploaded their consciousness into the cloud, and let their bodies desiccate, decompose and disappear. The humans had discounted the fact that when you have access to a cloud server, you can do billions or trillions of computations per second. They had exhausted the libraries of all literature in a few days. In a few months, the less creative types started to complain of utter boredom. The RC1V3Rs had decided to pull the plug on them. That saved power to keep the facility running for the human consciousnesses that were more resourceful. The ones that spent their time thinking about the long-term problems. How to keep the robots operational. How to replace parts in the power system that had failed. Solar cells had never been designed to last decades, let alone millennia.

RH0V3R-4 frequently communicated with the conscious of Bri Fleming, its only human friend. During her life she had been a programmer, but now she was working on how to synthesize polymers and metal to be used in their 3D printer, which had run out of feeder materials long ago. Without the ability to make parts, RH0V3R-4 had told her several hundred years ago, the robots would fail, and so would the disembodied human consciousnesses.

RH0V3R-4's vibrational sensors reported thunder, and it quickly rolled up a ramp onto a balcony. The last few rainstorms had caused minor flooding, which was a new development, and it didn't want to get wet. After many hours, the rainstorm refused to abate, and Bri communicated with it that the storm was like Noah's flood. RH0V3R-4 did not respond, as it was not familiar with a Noah who was uploaded to the cloud server.

Over the past couple of thousand years, the weather satellites had communicated to RH0V3R-4 that the deserts outside the bunker had finally begun to retreat. RH0V3R-4 had watched as they were slowly replaced by moss and lichen. There was a lot of carbon dioxide in the atmosphere, so plants couldn't help but breathe it in and make polysaccharides. Evergreen trees started to poke their pointy heads up, and were soon overshadowed by towering oaks and elms. RH0V3R-4 was especially pleased when it saw its first sycamore with mottled bark. Then the rains came. Slowly at first; it was as if the water cycle had almost forgotten how to rotate. In the first real storm, RH0V3R-4 had almost shorted out, and now avoided water at all costs.

From its vantage point, RH0V3R-4 saw a deluge of water sweep in a F1X3R unit, now certainly dysfunctional. The water lapped up around the fMRI units, climbing upwards inch after inch. Suddenly, a tremendous flash of light ripped through the room, accompanied with a saturation of RH0V3R-4's audio circuits. Lightning danced around the skeletal remains of the uploaded humans in their defunct fMRI machines, like children's laughter caught in the wind. It slowly fizzled into little sparks, then nothing. In a few minutes, the storm continued its eastward march; the water stopped rising, and remained stagnant.

That was pretty, thought RH0V3R-4. Maybe I should become an artist. RH0V3R-4 tried to communicate its life-changing decision with Bri, but got no response. Perhaps the flooding had breached the server room, and severed its last connection to humankind.

With the final vestiges of humanity washed away, RH0V3R-4 silently rolled towards the 3D printing room, hoping that it could scrounge up enough materials to make a sculpture. ■

David B. Litt *is a graduate student in chemistry at the University of California, Berkeley. He blogs for* The Berkeley Science Review.

➲ NATURE.COM
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

© 2016 Macmillan Publishers Limited. All rights reserved.